DOCUMENT RESUME

ED 026 718 24 EA 001 936

By-Shettel, Harris H.; And Others
Strategies for Determining Exhibit Effectiveness. Final Report.
American Institutes for Research, Pittsburgh, Pa.
Spons Agency-Office of Education (DHEW), Washington, D.C. Bureau of Research.
Report No-Project-V-001
Bureau No-BR-5-0254
Pub Date Apr 68
Contract-OEC-6-10-213
Note-244p.
EDRS Price MF-$1.00 HC-$12.30
Descriptors-*Changing Attitudes, Educational Objectives, *Evaluation Techniques, *Exhibits, *Interests, *Knowledge Level, Methodology, Theories, Time

This project was designed to develop research strategies and hypotheses for evaluating the effectiveness of exhibits. An exhibit on the role of the Federal Government in science and technology was used as the subject matter. Two basic groups of viewers were used, casual viewers and paid experimental viewers. Both were tested on knowledge gained and on interest and attitude levels. Control data were obtained for both groups. The paid subjects, obtained from high school, college, and adult populations, were divided by age, sex, education, and science and nonscience background. Time was the variable in the experiment, with one group of paid subjects having unlimited viewing time and one group being limited to one-half hour. Casual viewers were observed for comparison with the experimental groups and to determine the relative attractiveness of the exhibit elements. A mockup or small-scale replication of the exhibit was built to determine the feasibility of using such devices for prevalidation and experimental purposes. Results indicate that an exhibit can be analyzed using a wide variety of techniques and that studies of this type can make a major contribution to improved exhibit design and effectiveness. The mockup study also proved successful. A theory of exhibit effectiveness is described. (HW)

ED026718

BR 5-0254
PA-24

FINAL REPORT

Project No. V-011
Contract No. OE-6-10-213

STRATEGIES FOR DETERMINING EXHIBIT EFFECTIVENESS

April 1968

U.S. DEPARTMENT OF HEALTH,
EDUCATION, AND WELFARE
Office of Education
Bureau of Research

AMERICAN INSTITUTES FOR RESEARCH
Pittsburgh, Pennsylvania

EA 001 936

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

FINAL REPORT

Project No. V-011
Contract No. OE-6-10-213

STRATEGIES FOR DETERMINING EXHIBIT EFFECTIVENESS

Harris H. Shettel
Margaret Butcher
Timothy S. Cotton
Judi Northrup
Doris Clapp Slough

American Institutes for Research
Pittsburgh, Pennsylvania

April 1968

The research reported herein was performed pursuant to a contract with the Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U. S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

TABLE OF CONTENTS

PAGE

ACKNOWLEDGEMENTS viii

SUMMARY x

INTRODUCTION 1

METHODOLOGY 6

Definition of Major Experimental Variations 8
Experimental Design, Exhibit Testing Stage 12
Experimental Design, Mock-up Testing Stage 14
Empirical Effectiveness Variables -- Paper-and-Pencil Measures 15

Phase 1. Determining measurable objectives for the exhibit 16
Phase 2. Analyses of exhibit content 18
Phase 3. Development of background, interest, and attitude measures 25
Phase 4. Development of knowledge measures 27

Empirical Effectiveness Variables -- Nonpaper-and-Pencil Measures 33

Video tape analysis 33
Readability analysis 37
The measurement of typesize 37
Time data 39

Mock-up Testing 39

Phase 1. Mock-up validation testing 43
Phase 2. Mock-up variation testing 46

Summary 48

RESULTS 50

Biographical Data 50
Interest Results 50
Attitude Results 56
Open-End Concept, Open-End Knowledge, and Multiple-Choice Item Results 62
Exhibit-Only Results 64
Government Items Results 66
MAX Results on Exhibit Areas and Subareas as Determined by the Multiple-Choice Questionnaires 71
Total Knowledge Results 76
Results of Analyses of Variance on Selected Variables 83

TABLE OF CONTENTS (CONT'D.)

PAGE

Mock-up Validation Results 94
Mock-up Variation Results 102
Mock-up Variation Discussion 106
Results of the Video Tape Analysis 110
Results of Readability Analysis 117
Time Data Results 124
Determination of Attractiveness Scores 130
Casual Viewer Comments 133

DISCUSSION AND CONCLUSIONS 144

Objectives 144
Methods of Evaluation 146
Mock-up Validation 147
Audience Variables 148
Retention 149
Laboratory Approaches 151
The Role of Theory in Exhibit Research 152
A Three Factor Theory of Exhibit Effectiveness 153

REFERENCES 160

ERIC REPORT RESUME 230

LIST OF FIGURES

Figure 1. Three categories of variables considered in this study. 7
Figure 2. Physical layout of Chicago exhibit. 19
Figure 3. View of judging area showing judge, Ampex Video Tape System, "Fish Sounds" placard, television monitor, Esterline-Angus Event Recorder, and an additional television monitor. 36
Figure 4. Photographs of exhibit subarea, DNA, and corresponding DNA mock-up panel. 41
Figure 5. Mock-up panel of introductory display, Man and His Earth 42
Figure 6. Physical layout of the mock-up validation. 44
Figure 7. Photographs of study subjects viewing mock-up introductory and subarea panels. 45
Figure 8. Chicago casual viewers' mean interest ranks of exhibit areas. 53
Figure 9. Chicago exhibit area mean interest rankings for MIN, MAX, and CONTROL within each age group. 55

PROCESSING CONTROL RECORD

for

MUST and MUST - EXPEDITE DOCUMENTS

ERIC FACILITY

Document Identification Number BR-5-0254

Date Received from Central ERIC DEC 10 1968

Date Shipped to Clearinghouse DEC 16 1968

ERIC CLEARINGHOUSE

Date Received Jan 06 1969

Clearinghouse Accession Number EA 001 936

Date Shipped 4-18-69

ERIC FACILITY

Date Received April 21, 1969

ED Accession Number ________

Date Shipped to EDRS ________

LIST OF FIGURES (CONT'D.)

PAGE

Figure 10. Overall attitude percent correct for Chicago and Los Angeles subjects. 57
Figure 11. Summary of attitude percent correct for exhibit objectives. 60
Figure 12. Percent correct, attitude scores, for Chicago science and nonscience study groups. 61
Figure 13. Percent correct for Chicago groups on three recall Government items. 67
Figure 14. Percent correct on Government items. 68
Figure 15A. Chicago total MAX group percent correct on multiple-choice items for areas and subareas. 72
Figure 15B. Chicago total MAX group percent correct on multiple-choice items for areas and subareas. 74
Figure 15C. Chicago total MAX group percent correct on multiple-choice items for areas and subareas. 75
Figure 16. Total knowledge percent correct learning continuum for Chicago experimental groups 77
Figure 17. Summary of science/nonscience groups on total knowledge percent correct. 80
Figure 18. Summary of total knowledge percent correct experimental viewing conditions within age/education groupings. 82
Figure 19. Idealized distributions of interest scores for the four groups. 89
Figure 20. Idealized distributions of open-end concept scores for the four groups. 90
Figure 21. Idealized distributions of open-end knowledge scores for the four groups. 91
Figure 22. Idealized distributions of multiple-choice scores for the four groups. 92
Figure 23. Idealized distributions of the exhibit-only scores for the four groups. 93
Figure 24. Mock-up and Chicago mean attitude scores for each exhibit objective. 96
Figure 25. Mean interest ranks for MIN, MAX, and CONTROL high school subjects in mock-up validation. 97
Figure 26. Relationship between actual exhibit and exhibit mock-up in terms of uncorrected mean scores for 14 subareas. 99
Figure 27. Relationship between actual exhibit and exhibit mock-up in terms of corrected mean scores for 14 subareas. 100
Figure 28. Total knowledge scores (percent correct) of science and nonscience subjects by group in mock-up variation study. 107
Figure 29. Sample Esterline-Angus Tape, video tape analysis. 112

LIST OF FIGURES (CONT'D.)

PAGE

Figure 30. Video tape analysis, "Fish Sounds" exhibit. 113
Figure 31. Video tape analysis, "Desalination" exhibit. 115
Figure 32. Summary of Chicago casual viewer time data for each exhibit area. 125
Figure 33. Exhibit attracting power (in percent) and holding power (in seconds) for each subarea. 126

LIST OF TABLES

Table 1. Design Specifications; Exhibit Phase #1 12
Table 2. Design Specifications; Exhibit Phase #2 13
Table 3. Design Specifications; Mock-up Phase #1 14
Table 4. Design Specifications; Mock-up Phase #2 15
Table 5. Original and Translated Objectives for Vision of Man Exhibit 17
Table 6. Label and Description of Each Exhibit Area and Subarea 20
Table 7. Relationship Between Exhibit Objectives and Attitude Items in the Questionnaire 26
Table 8. Chicago Word Count, Area Percentages, and Number of Items for Multiple-Choice and Exhibit-Only 30
Table 9. Number of Words, Reading Grade Levels, and Average Typesize in Exhibit Areas and Subareas 38
Table 10. Summary Table of Total Number of Words and Text Grade Levels for Full and Skeleton Texts Used in the Mock-up Variation Testing Phase 47
Table 11. Answers of Chicago Posttest Groups to Interest Question: "Did the Exhibit Stimulate your Interest in a Particular Topic?" 51
Table 12. Chicago High School MIN and MAX Groups Percent Correct on Comparable Concept, Knowledge, and Multiple-Choice Items 63
Table 13. Percentage Scores on Exhibit-Only Questionnaire by Experimental Group and by Educational Level 65
Table 14. Gain Scores over Control Groups by Experimental Group and by Educational Level, Exhibit-Only Questionnaire 65
Table 15. Effects of Experimental Conditions (Based on Experimental Subjects) 84
Table 16. Age/Education Differences (Based on Experimental Subjects) 84
Table 17. Sex Differences (Based on Experimental Subjects) 85

LIST OF TABLES (CONT'D.)

		PAGE
Table 18.	Science Background Differences (Based on Experimental Subjects)	85
Table 19.	Effects of Socioeconomic Status (Based on Experimental High School Subjects)	86
Table 20.	A Comparison of Casual Viewers with Experimental Viewers and Nonviewers (Based on All Subjects)	86
Table 21.	Mock-up Validation Results (Based on Experimental and Mock-up Subjects)	95
Table 22.	Mock-up Variation Percentage Scores by Group and by Test	103
Table 23.	Mock-up Variation Percentage Scores by Group and by Subject Matter	103
Table 24.	Mock-up Variation Total Knowledge Percentage Scores According to the Three Variables of the Study	104
Table 25.	Analysis of Variance of Three Mock-up Variation Variables (Dependent Variable = Total Knowledge Scores)	105
Table 26.	Correlation between IQ and Total Knowledge for Each Group in Mock-up Variation Study	105
Table 27.	Rank Order Correlations and Significance Levels for Judgmental Ratings for Two Exhibit Displays	111
Table 28.	Amount Learned on Two Items Based on Two Different Kinds of Displays	114
Table 29.	Total Knowledge Scores of Mock-up Variation Groups by Readability	118
Table 30.	Flesch Grade Level Computation for the DNA Subarea by Paragraph	122
Table 31.	Gain Scores for Low and High Items Drawn from DNA Subarea by Grade Level of Source Paragraph	123
Table 32.	Number of Casual Viewers Stopping as Predicted by Formula and Actual Number Stopping for Each Subarea (N=60)	132
Table 33.	Percent of Casual Viewer Comments Categorized as Positive, Basically Positive, Negative, and Neutral	133
Table 34.	Percentage of Casual Viewers Who Mentioned Five Most Frequent Categories	136
Table 35.	Percentage of Casual Viewer Comments Classified as Level-Positive, Level-Negative, Organization-Positive, Organization-Negative	136
Table 36.	Frequency of Mention of Various Exhibits by Casual Viewer	137
Table 37.	Main Characteristics of Los Angeles and Chicago Casual Viewer and Study Groups (All Experimental Groups Combined)	208

LIST OF APPENDIXES

	PAGE
Appendix A - Story Line for Exhibit on Federal Science and Engineering for Museum of History and Technology Exhibit Objectives	163
Appendix B - Fact Sheet on the Federal Science and Engineering Exhibit -- The Vision of Man	176
Appendix C - Physical Layout of Los Angeles Exhibit	185
Appendix D - Posttest Background Questionnaire	186
Appendix E - Post-Interest Index	187
Appendix F - Attitude Questions	188
Appendix G - Open-end Concept Questions	192
Appendix H - Open-end Knowledge Questions	193
Appendix I - Multiple-Choice Questions	196
Appendix J - Exhibit-Only Questions	204
Appendix K - Physical Layout of the Mock-up Variation	206
Appendix L - Summary of Background Questionnaire Results	207
Appendix M - Individual Cell Frequencies and Means for the Analyses of Variance	211

ACKNOWLEDGEMENTS

The original project staff consisted of Mr. Harris H. Shettel as Principal Investigator, Mrs. Doris Clapp Slough as Project Director, and Mrs. Margaret Butcher as Assistant Project Director. Mrs. Slough remained with the project throughout all of the exhibit testing phases, but left the project before the data analysis had been completed and before the final report had been prepared. Upon her resignation, Mr. Shettel and Mrs. Butcher shared Project Director responsibilities for the remainder of the project. Two members of the Institute research staff were added to the project specifically to assist in the preparation of this report. They are Mr. Timothy S. Cotton, who completed the data analysis, and Miss Judi Northrup, who wrote several sections of the report.

Instrumental in the successful completion of the testing phases of this study was the excellent cooperation and assistance provided by museum personnel who permitted testing of museum visitors, arranged for testing areas to be set aside within the museum, and extended many courtesies to the project staff. They are:

Mr. Henry Sturr, Director, California Museum of Science and Industry, Los Angeles, California.

Mr. William C. O'Donnell, Director of Education, California Museum of Science and Industry, Los Angeles, California.

Mr. Daniel M. McMasters, Director, Museum of Science and Industry, Chicago, Illinois.

Mr. Vernon J. Pietz, Architect - Manager of Exhibits, Museum of Science and Industry, Chicago, Illinois.

Dr. O. S. Williams, Educational Supervisor, Museum of Science and Industry, Chicago, Illinois.

The various and complex testing phases of this study ran smoothly due to the efforts of:

Mr. Allen J. Klaus, supervisor of casual viewer testing in Los Angeles and Chicago.

Mrs. Julia Ashenhurst, Conference Coordinator, Center for Continuing Education, University of Chicago.

Mr. and Mrs. Edward Lotko, Lions Club, Chicago, Illinois.

Dr. Mark Provus, Director of Research, Board of Education, Pittsburgh, Pennsylvania.

Technical consultants made important contributions to the sensitivity and accuracy of the various questionnaires developed in this study by reviewing the exhibit text and critiquing individual test items in their field of special interest. These consultants were:

Dr. Ray C. Hackman, Psychological Service of Pittsburgh -- Attitude Measures.

Dr. Allen Janis, University of Pittsburgh -- Physics.

Dr. Erston V. Miller, University of Pittsburgh -- Biology.

Dr. N. E. Wagman, University of Pittsburgh -- Astronomy.

Dr. Harry J. Werner, University of Pittsburgh -- Earth and Planetary Sciences.

Miss Geraldine Spaulding -- Test Construction Consultant -- Reviewed all questionnaires.

The video taping of casual viewers was accomplished with the assistance of Mr. John Fetter, whose equipment and expertise made it possible. Mr. David Elder was largely responsible for the excellent design and construction of the exhibit mock-up.

And finally, a vote of thanks to all of the many adult, high school, and college subjects who participated in the experimental phases of this work. They were asked to do many strange things, and they almost always did them.

SUMMARY

The purpose of this project is to initiate the systematic development of research strategies and testable hypotheses that will make it possible to better evaluate the effectiveness of scientific and technical exhibits, particularly those designed to reach educational objectives. Exhibit effectiveness is conceptualized as a measurable change in viewer behavior produced by the exhibit, and consistent with the stated aims or objectives of the exhibit. The development and refinement of this approach is seen as a necessary prerequisite to the collection of valid data relating not only to the evaluation of existing exhibits, but also to the development of a body of knowledge that will meaningfully relate effectiveness variables to exhibit design variables.

Expected Contribution to Education. Better utilization of exhibits must be predicated upon a sound methodology for determining exhibit effectiveness. Improving this methodology will provide exhibit sponsors and designers with a more rational basis for making decisions regarding the resources they should devote to exhibits. The development of objective criteria of effectiveness would also make it possible to choose the most appropriate kind of exhibit from among various types to accomplish given objectives.

Methodology. A complex, modern exhibit, designed to impart knowledge about the role of the Federal Government in science and technology, and to develop a favorable attitude in young people toward this role, was evaluated in two geographical locations, Los Angeles and Chicago.

A variety of research techniques and experimental variables were evaluated. A basic aim of the study was to find out how much could be learned from an exhibit given "maximum" time and motivation. Thus, a maximum index of effectiveness measure (MAX) was established to determine changes occuring in the levels of knowledge, interests, and attitudes among viewers spending unlimited time in an exhibit under instructions to learn "as much as possible." A minimum index of effectiveness measure (MIN) was established to determine knowledge, interest, and attitudinal changes occuring among viewers spending a "minimum" amount of time (1/2 hour) in the exhibit. A CONTROL group, made up of people who had not seen the exhibit, established base scores on all of the various test measures. Those tested within these experimental time groupings were volunteer subjects, paid for their services. These subjects were divided by age, education, and sex. The three basic age and education groups were high school students, college students, and adults. Since the exhibit dealt with recent advances in science and technology, each group of subjects was further divided into science/nonscience groups on the basis of educational background and course work.

Casual viewers represented a second major category of subjects. The casual viewer data make possible the comparison of "real" museum visitors with the various experimental groups just described. In addition to the basic knowledge, interest, and attitude data, observations of crowd flow and time spent at various displays were recorded and analyzed for casual viewers. They were also asked to comment on the exhibit; these statements were tape recorded and later analyzed.

Several innovative techniques were developed in an effort to investigate various exhibit design variables that may contribute to the overall effectiveness of an exhibit. Casual viewer time data was used to establish an "attractiveness rating" for various exhibit subareas. In addition, the relative "attracting power" of individual design elements (models, pictures, signs, etc.) within an integral display unit was investigated by means of a video taping technique in which the behavior of casual viewers was recorded on tape for later analysis. A readability analysis was performed on the exhibit textual materials in order to determine the reading difficulty levels of individual subareas. An exhibit mock-up, consisting of a small-scale simulation of the entire exhibit, was developed from the pictures and text of the original exhibit. The mock-up was validated against the actual exhibit by replicating the entire experimental design with high school students. A portion of the mock-up was used to experimentally explore several important design parameters, including the use of sound, the use of illustrations, and reading difficulty of the written text.

An important prerequisite of the above data collection procedures was the analysis and refinement of the objectives of the exhibit, and the development of comprehensive testing devices capable of measuring these objectives. Test formats used for knowledge items included multiple-choice, open-end knowledge, open-end concept, and exhibit-only. Attitude and interest measures also utilized a variety of formats. Biographical information was collected from all subjects in a background questionnaire incorporating items on age, education, science background, and interest in science.

Results. The amount of viewing time and the motivation of the viewer definitely influences the amount of knowledge gained from the exhibit. Combined total knowledge scores, derived from the summation of the four individual knowledge scores, show that the MAX group (unlimited viewing time) always achieved higher scores than the MIN group and that the MIN group scores were always higher than the CONTROL group scores. The casual viewer (museum visitors) pretest scores are similar to the casual viewer posttest scores, and actually lower than the CONTROL group scores. Thus, the casual viewer group, as a whole, learned very little from the exhibit as measured by the tests used. In general, college students learned considerably more from the exhibit than either of the other two age groups. In turn, high school subjects performed at a higher level than adult subjects. There was a significant difference between science and nonscience subjects on this combined measure, with science subjects attaining consistently higher total knowledge scores.

Results indicate that interest levels can be influenced by viewing an exhibit, although the direction and extent of the changes were not always comparable between the various groups. The CONTROL group data established the initial expectations of interest for the viewing groups. The MIN and MAX viewing groups showed widely divergent interest patterns in comparison to the CONTROL group data. The pretest and posttest casual viewer groups also showed divergent interest patterns. In several instances, high pre-interest patterns became low post-interest patterns, indicating that the exhibit failed to sustain the initial level of interest in these particular areas. The opposite effect was also noted.

While the knowledge and interest results show significant differences between viewing and nonviewing groups, the attitude data generally do not. All experimental time groups attained relatively equal scores as did the casual viewer groups; there were no significant differences between CONTROL, MIN, and MAX groups. These results are consistent with the findings of other exhibit studies, i.e., attitudes seem not to be influenced by short-term exposure to an exhibit. However, lack of sensitivity of the measuring instrument and the possibility of long-term "sleeper" effects should not be overlooked.

In addition to test results, several other types of casual viewer data were collected and analyzed. The amount of time an average casual viewer spent at an exhibit subarea (41 subareas in total) equaled 20 seconds and the total time in the exhibit area equaled 14 minutes. These results indicate that the exhibit could not hold the average casual viewer's attention for more than a short period of time. Sixty casual viewers, without their knowledge, were followed throughout the exhibit and the displays where they stopped were noted. Such information can be used as a diagnostic tool in determining the relative "attracting power" of various displays, i.e., the percentage of casual viewers who stop at each display. Recordings were made of casual viewer answers and comments to the question, "What will you tell your friends and family about this exhibit?" Many useful suggestions were obtained in this manner, most of them having to do with the complexity of the exhibit.

Another technique for the measurement of "attracting power" involved the use of a multiple regression equation based on: 1) the number of subjects actually stopping at each subarea, 2) the number of static models, and 3) the number of dynamic models. This equation resulted in an "attractiveness rating" for each subarea. The formula generated by these data must be validated in other studies before being accepted as an "established" exhibit effectiveness technique.

Relative "attracting power" among exhibit subareas was also measured by video taping casual viewers as they viewed the display elements contained in two exhibit subareas. The frequency at which viewers looked at particular elements was estimated independently by three judges. Intercorrelations showed high reliability among the raters. The results indicate that there are large differences in attracting power among display elements. In general, dynamic (moving) models and their associated text were looked at by casual viewers much more frequently than static (immobile) models or pictures, and their text materials. The video taping technique could prove to be a valuable diagnostic tool in the evaluation of an exhibit's effectiveness and in the collection of basic data relating to exhibit design variables.

The amount of information that can be gained from an exhibit may be influenced by the reading difficulty of the exhibit text. A readability analysis, using the 1948 Revised Flesch Readability Formula, was performed on all the exhibit texts. Reading grade levels for exhibit subareas ranged from 6.8 to 14.8 years. Efforts to measure the influence of readability on knowledge gained were largely inconclusive due to a number of uncontrolled factors (e.g., placement of text, size of type, lighting).

The exhibit mock-up was built for two purposes: to explore the feasibility of constructing a representation of an exhibit for pre-validation purposes and to perform experimental variations upon exhibit design variables within the more flexible mock-up structure. The mock-up was validated by comparing exhibit and mock-up test scores across all high school experimental groups. The results indicate that mock-up subjects performed as well as exhibit subjects on the various tests. Once the mock-up had been validated, the following exhibit design variables were manipulated within the mock-up design: 1) hearing versus reading textual material, 2) visuals versus no visuals, and 3) full textual material versus skeleton textual material (a shorter, lower grade level text). An analysis of variance performed on these data indicates a significant difference favoring the group reading the text over the group hearing the text. There were no significant differences among the other design variables.

The demonstration of mock-ups as valid representations of exhibits could make a worthwhile contribution to exhibit effectiveness if they were used with a sample of the intended audience prior to the construction of the actual exhibit. Since the mock-up approach lends itself to design variations, changes in the planned exhibit could be made before costly "errors" were built into the final product.

The report concludes with a discussion of a three-factor theory of exhibit effectiveness. The theory attempts to account for the three areas found to be most significant in the present study: 1) initial attracting power, 2) holding power, and 3) teaching effectiveness. Hypotheses related to these three factors are suggested for investigation in future exhibit research studies.

INTRODUCTION

The current use of scientific and technical exhibits as educational tools is both widespread and costly. However, evaluations of their demonstrated effectiveness are infrequent and there are no generally accepted standards for performing such evaluations. The need for developing standardized techniques of evaluation is being increasingly recognized, not only to assist in determining the effectiveness of existing exhibits, but also to serve as guidelines in the design of new exhibits.

The techniques of evaluation most typically used tend to fall into two broad categories: 1) an appraisal by exhibit "experts" or, 2) ad hoc/empirical measures of a particular exhibit's effectiveness, often based on peripheral indices such as attendance figures.

"Expert" ratings of exhibits have been shown in at least one instance to be unreliable. In a previous study by the American Institutes for Research, a rating scale was developed that reflected the criteria for exhibit effectiveness as stated in the relevant literature. Using this scale, it was found that the interjudge correlations between experts' ratings of specific features of several exhibits were low, the median being .24 (19). Another study (3) examined the reliability of overall ratings of exhibits, based on a combination of criteria: 1) clarity of presentation, 2) general attractiveness, 3) integration of panels, and 4) value of the information portrayed. The reliability of ratings based on these criteria was also low.

In short, it appears that there is a lack of agreement among experts in the exhibit field on the interpretation of the various rating criteria that have traditionally been used to evaluate exhibits, and thus it could not be recommended that such criteria be used as a valid basis for determining actual exhibit effectiveness.

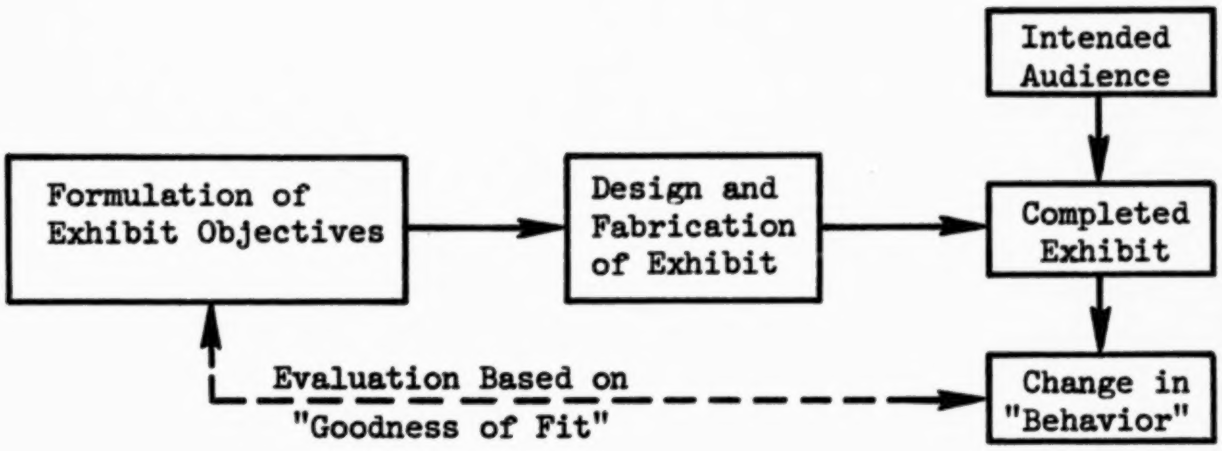
The second technique, i.e., using ad hoc/empirical measures of an exhibit's effectiveness, tends to serve only as documentation of overt success or failure, and does not provide data that can be used as a basis for improvement in the design (and effectiveness) of exhibits in general. It is true that this approach, even though nonsystematic, might lead to improved designs over a period of years. However, without objective criteria for comparing results, one could only predict slow and halting progress, comparable to the very gradual improvement in textbooks, films, and other media of communication and education prior to the fairly recent introduction of more analytic and systematic studies of their effectiveness. The heterogeneity of exhibit viewers and the complexity of exhibits undoubtedly present unique problems. But, the increased use of exhibits, combined with their rising costs, makes it imperative that steps be taken to develop rational approaches to the assessment of effectiveness.

Before describing the study itself, it would be well to define two key terms: "exhibit" and "effectiveness." In its broadest sense, an exhibit is a display for public inspection. Specific exhibits differ in major design characteristics, such as size, media used, materials of construction, the number and kinds of models used, color and lighting characteristics, etc. These internal variables are manipulated to account for external variables such as space available, subject matter, audience characteristics, viewing time, location relative to other exhibits, and, of course, cost. The end result of the "mating" of these two sets of variables is an exhibit that meets (hopefully) certain needs and/or objectives.

The meaning of the term "effectiveness" also varies, interacting with the purpose of the exhibit. Thus, a crucial distinction between an advertising exhibit designed for a trade fair and a didactic exhibit designed for a children's museum would be their goal or purpose. The advertising exhibit would be designed to "stimulate interest," whereas the purpose of the educational exhibit would be to "achieve understanding of relationships," or to "increase knowledge" (4 & 5). In one case, attendance figures may suffice as an index of effectiveness, whereas in the other case, more elaborate testing techniques would be required to demonstrate effectiveness. Furthermore, the criterion of success for even one major class of exhibits, e.g., scientific and technical, may vary at the discretion of the individuals or groups connected with the exhibits. Thus, it is necessary to consider effectiveness in the context of different exhibit objectives.

Even if a systematic basis for determining effectiveness were developed as a result of successive stages of research, it would undoubtedly have to be revised from time to time simply to account for changes in the concept of an exhibit and corresponding changes in specified objectives. Changes of this sort have already occurred in the evolution of science museums from their original use as storehouses for collections of historical artifacts with little or no explanation of their function to an increasing emphasis on the teaching of scientific knowledge and principles, with ample context and operational models to achieve the desired educational objectives. Thus, there has been a general shift from "cataloging" to "instruction" (11).

At this point it is evident, even before further analysis, that establishing objective criteria for exhibit effectiveness will require an analysis of the relationships between three elements: exhibit objectives, variables intrinsic to exhibit design, and variables extrinsic to exhibit design. These three elements are contained in the meaning of exhibit effectiveness as conceptualized in this study: *Exhibit effectiveness is demonstrated on the basis of a measurable change in the behavior of the intended audience, produced by the exhibit, and consistent with the stated aims or objectives of the exhibit.* The term "behavior" is meant to include a range of overt acts; the ones most relevant for a particular exhibit would depend on the purpose of the exhibit and the availability of the behavior for measurement. Because terminal, or criterion, behavior is frequently difficult or impossible to observe, it is often necessary to use abstractions of that behavior, such as answering test items, stating an opinion, or expressing an attitude. Such abstractions must be carefully prepared lest they distort the original aim of the exhibit, or even replace it entirely. When this happens (as it does, for example, when attendance figures or expressions of audience interest are used to indicate the effectiveness of an exhibit designed to impart factual information) the meaning of effectiveness, as conceptualized here, is lost entirely. This conceptualization is shown graphically in the following diagram:


Intended Audience
Formulation of Exhibit Objectives
Design and Fabrication of Exhibit
Completed Exhibit
Evaluation Based on "Goodness of Fit"
Change in "Behavior"

Purpose of Study

The ultimate goal of a program of exhibit effectiveness research would be to establish objective criteria for measuring the effectiveness of scientific and technical exhibits designed to reach educational objectives. Such criteria, once established, would be applicable both to the evaluation of existing exhibits and to the design of proposed exhibits.

To be maximally useful, criteria that are established for use in the design and development of *new exhibits* should not require the collection of data from viewers of the completed exhibit. Therefore, such criteria would not actually measure any change in viewer behavior in accord with the exhibit's objectives, but they would have predictive validity for such changes. Criteria suggested for use in the evaluation of the *completed* exhibit may include both the design criteria and any empirical measures that are valid and efficient, i.e., requiring relatively little time and cost to develop and apply. Empirical measures that are recommended would measure either a change in viewer behavior corresponding to the stated aim or objective of the exhibit or any "static" behaviors that have predictive validity for behavior changes in accord with the exhibit's objectives, e.g., "attending" behaviors such as average viewing time or percent of viewers stopping at a particular display.

It is obviously not possible to achieve all of these goals in one study. Before any measures of effectiveness can be established that are not based on data collected from viewers of the finished exhibit, an independent set of measuring techniques must be developed that can be used now to set up testable hypotheses regarding objective criteria, and used again in subsequent stages of research to determine the final set of criteria. Thus, the scope of this study is limited to exploring evaluation methodologies and generating testable hypotheses; it is not intended to arrive at any final conclusions in terms of the effectiveness of specific design parameters, location parameters, etc. Work in related fields such as films, educational television, and programmed instruction suggests that such a goal is clearly a long-range one.

At this initial stage of research any set of measures of effectiveness was necessarily empirical. For this purpose, a particular exhibit was designated as the basis for generating the required set of measures and for collecting data from exhibit viewers. By varying the characteristics of the viewers and the viewing conditions, and by describing and measuring certain design variables, it was possible to collect data that could be used as a basis for recommending various empirical measures for further stages of research and simultaneously, for the generation of testable hypotheses regarding objective criteria. In addition, it was possible to set up a small-scale experimental study using a simulation, or mock-up, of the exhibit, along with variations, to explore some of the hypotheses generated on the basis of the analysis of the data from the "real" exhibit.

The "subject matter" for this study was a large, Federally sponsored, scientific and technical exhibit entitled "The Vision of Man." The exhibit was designed to cover the more important, nonmilitary, scientific and technical programs that have been, and are being, conducted by the Federal Government, often in conjunction with private industry, and the resulting accomplishments of such programs. It was designed to appeal to a wide audience, but to be especially attractive to the high school age bracket who may be influenced in career choice (science) and selection of employer (Federal Government) by the exhibit. The exhibit was coordinated by the U. S. Civil Service Commission, designed by Herbst-Lazar, Chicago, Illinois, and constructed by General Exhibits, Inc., Philadelphia, Pennsylvania. It first appeared in the Smithsonian's Museum of History and Technology after which it went to the Federal Pavilion at the 1964 New York Fair. It then moved to Los Angeles (California Museum of Science and Industry), where part of this study was carried out, and later to Chicago (Museum of Science and Industry), where additional testing was accomplished. The exhibit itself was large (5,000 square feet) and complex (over 40 individual displays) and was a self-contained major attraction in the museum display area. Considerable local publicity attended the exhibit, particularly when it was initially opened to the public.

The scope and complexity of the Vision of Man Exhibit presented problems in evaluation that would not exist in an exhibit with more limited aspirations. On the other hand, this complexity was an advantage in the context of an exploratory methodological study because it did allow the staff to explore a wide variety of design elements utilizing current exhibit technology, and attempt to assess goals that clearly included attitude, interest, and knowledge components. In short, the exhibit could be considered to represent a modern, sophisticated "state-of-the-art" display, and thus be an appropriate "culture" in which to test the "potency" of a variety of effectiveness measures.

Unavoidably, however, this study is colored and limited by the characteristics of the particular exhibit selected for examination. But this limitation has a greater impact on the specific results obtained than it does on the methods used to obtain those results. Thus, the reader is asked to remember (and will be reminded) that it is not the purpose of this study to evaluate the Vision of Man Exhibit; the results per se are of little consequence. What is of consequence and what is being evaluated are the methods used to perform the evaluation.

The reader may detect a logical tautology in this conceptualization, i.e., you cannot use an unknown quantity (the Vision of Man Exhibit) as the basis for establishing "universal" standards of measurement. But most psychological measures face this problem, including those designed to evaluate intelligence, aptitude, interest, and even knowledge. There is no reasonable alternative but to work toward standards of measurement through an iterative process of successive approximations. An unreasonable alternative is to fall back on measures that have high reliability and perhaps face validity, but little or no actual validity. Attendance figures, square footage of exhibit space, cost, number of displays, etc. are examples of such measures. The present study rejects these approaches and attempts to deal with substantive issues, even though these issues present formidable difficulties in design and analysis. Furthermore, the study described here makes an effort to explore as many approaches to exhibit evaluation as possible within the limitations of time and funds available. In fact, several evaluation techniques were added on to the already complex design of the study simply because they were felt to be potentially useful and productive (which they were). Favoring scope rather than depth has an advantage in an exploratory study since it provides a greater number of "leads" for additional research. In this way, progress can be made toward a better understanding of the exhibit medium in all of its complexity.

METHODOLOGY

A scientific, technical exhibit can be thought of conceptually in several ways, depending on one's professional orientation. A learning psychologist might see it in terms of a stimulus and response paradigm, and may further conceive of the exhibit as being concerned with items similar to those considered relevant to modern educational (programmed) materials, e.g., behavioral objectives, size of step, sequence of steps, response elements, etc. A communications theorist might see an exhibit in terms of "sender" (sponsors, goals, and objectives), "channel" (the exhibit itself), and "receiver" (the viewer). The present study takes an essentially empirical, operational approach, but one grounded in modern educational practices.

This exhibit is seen to consist of three sets, or clusters, of variables. For purposes of identification in this report, these three types of variables will be called, respectively: 1) exhibit design variables, 2) exhibit viewer variables, and 3) exhibit effectiveness variables. The first could be considered as independent or experimental variables, the second as control variables, and the last as dependent or criterion variables. The hypotheses to be investigated should attempt to identify which are the most important variables of each type, their relationship with one another, and the best way in which to measure them.

The variables of each of the three types indicated above that were selected for examination in this study are shown in Figure 1. The variables are indicated on a three-dimensional figure in order to represent their possible interaction.

In general terms, the exhibit design variables that were evaluated were physical characteristics of the exhibit itself (e.g., readability level of labels, internal location of the various parts of the exhibit, etc.); the exhibit viewer variables included audience characteristics (e.g., age/educational level) and viewing conditions (e.g., extent to which visit is voluntary and the viewing time); the effectiveness variables included observations of viewers in order to arrive at measures of attraction and holding power, and various test batteries designed to measure knowledge, attitude, interest, etc.

The study involved two major research stages, the Exhibit Testing Stage and the Mock-up Testing Stage. The Exhibit Testing Stage is so called because the basis for testing at this point was the exhibit itself; during the Mock-up Testing Stage, the basis for testing was a small-scale simulation, or mock-up, of the exhibit.

The Exhibit Testing Stage was used to set up, test and revise the measures of empirical effectiveness; in addition, data were collected regarding the relations among the different types of variables. During the Exhibit Testing Stage, experimental variations could not be made in exhibit design variables, but experimental variations could be made in some of the exhibit viewer variables. The variations that were

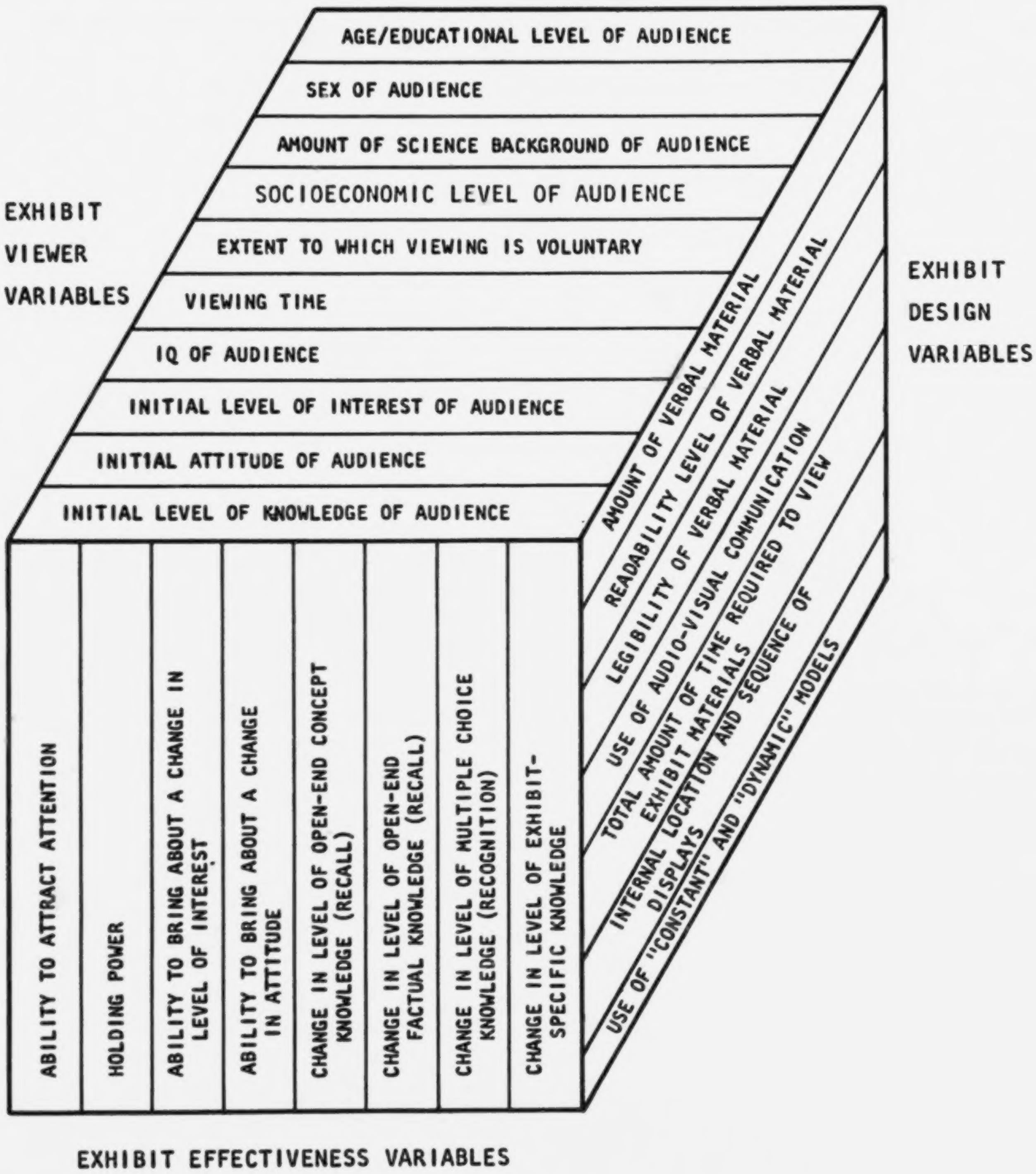

AGE/EDUCATIONAL LEVEL OF AUDIENCE
SEX OF AUDIENCE
AMOUNT OF SCIENCE BACKGROUND OF AUDIENCE
SOCIOECONOMIC LEVEL OF AUDIENCE
EXTENT TO WHICH VIEWING IS VOLUNTARY
VIEWING TIME
IQ OF AUDIENCE
INITIAL LEVEL OF INTEREST OF AUDIENCE
INITIAL ATTITUDE OF AUDIENCE
INITIAL LEVEL OF KNOWLEDGE OF AUDIENCE
EXHIBIT VIEWER VARIABLES
ABILITY TO ATTRACT ATTENTION
HOLDING POWER
ABILITY TO BRING ABOUT A CHANGE IN LEVEL OF INTEREST
ABILITY TO BRING ABOUT A CHANGE IN ATTITUDE
CHANGE IN LEVEL OF OPEN-END CONCEPT KNOWLEDGE (RECALL)
CHANGE IN LEVEL OF OPEN-END FACTUAL KNOWLEDGE (RECALL)
CHANGE IN LEVEL OF MULTIPLE CHOICE KNOWLEDGE (RECOGNITION)
CHANGE IN LEVEL OF EXHIBIT-SPECIFIC KNOWLEDGE
AMOUNT OF VERBAL MATERIAL
READABILITY LEVEL OF VERBAL MATERIAL
LEGIBILITY OF VERBAL MATERIAL
USE OF AUDIO-VISUAL COMMUNICATION
TOTAL AMOUNT OF TIME REQUIRED TO VIEW EXHIBIT MATERIALS
INTERNAL LOCATION AND SEQUENCE OF DISPLAYS
USE OF "CONSTANT" AND "DYNAMIC" MODELS
EXHIBIT DESIGN VARIABLES
EXHIBIT EFFECTIVENESS VARIABLES

Figure 1. Three categories of variables considered in this study.

experimentally manipulated related to age, educational and socioeconomic level, sex, amount of science background, the extent to which the exhibit viewing was voluntary, and viewing time.

The Mock-up Testing Stage was used in order to investigate, and generate hypotheses regarding, the effects of experimental variations in several exhibit design variables whose manipulation in the exhibit itself was not possible. The experimental variations in the mock-up related to the simultaneous condensation and simplification of text materials, the amount of visual illustration, and the use of audio devices to present (and replace) the textual portion of the exhibit.

Definition of Major Experimental Variations

Since this was an exploratory study, it was considered proper to investigate the effects of those elements expected to make the most difference; those that appeared to be significant would be suggested for further, and more refined, analysis.

Nature of the visit. Two basic audience variations were used: unsolicited visitors to the museum selected on a random basis, called "casual viewers," and persons solicited and paid to participate in the study, called "study groups." The "study groups" are described in detail below under the variable of viewing time. First, however, the approach taken to the "real" audience of the exhibit, the casual viewer, will be described.

Several types of data were collected on this group at both of the exhibit testing sites. Prior to the actual testing periods, the project staff selected one field supervisor and three interviewers to handle the responsibility for all casual viewer data collection while the staff was concurrently testing the paid study and control subjects.

During the Los Angeles field tryouts, an effort was made to randomly select casual viewers to participate in the study. However, this selection process was requiring too much time due to the fact that there were relatively few exhibit visitors and consequently few participants. Since the casual viewer data was considered vital to the success of the study, the "rules" were modified so that any visitor who looked mildly interested in the entrance area of the Vision of Man was asked to participate and, if he agreed, became part of the pretest casual viewer group. Posttest casual viewers were asked to participate at the exhibit exit only, and did not know they were part of the study until they had completed their visit through the exhibit.

The field supervisor who was responsible for all of the solicitation asked for pretest and posttest participation at different time intervals, i.e., two hours of precasual viewer testing would be followed by two hours of postcasual viewer testing. This was done to simplify paper work, record-keeping activities, etc. Children of elementary school age were not asked to participate. Potential pretest subjects were told that "the Office of Education wants to find out more about the people who visit this exhibit," while potential posttest subjects were told that "Office of Education wants to find out what people thought of this exhibit." Both groups were asked to help by "completing a 15-minute questionnaire."

The level of acceptance by casual viewers was determined during the Chicago testing. Out of a total of 604 people who were asked to participate, 368 accepted, while 236 refused. The overall rate of acceptance was thus 61 percent. These data were not collected in Los Angeles.

In order to maintain an adequate acceptance ratio, it was felt necessary to limit the casual viewer testing time to a 15-minute period. This was considerably shorter than the test time required for study subjects (approximately 2-1/2 hours). Each casual viewer was first asked to fill out the background questionnaire and interest index. The attitude and knowledge items from the large test battery were randomly divided into eight subsets. (Specific knowledge items asking about the Federal Government were purposely distributed into each subset.) Each casual viewer was given one of these subsets. Thus, a score for a complete set of knowledge items represents the summation of eight individual casual viewers' scores on the different subsets. While this arrangement presents some difficulties with data analysis procedures, it seemed preferable to the alternatives:

1. Pay the casual viewer.
2. Pick casual viewers who would take two hours to answer the questionnaire.
3. Develop a special test requiring only 15 minutes.

One and two would provide highly biased data on the nature of the true population of casual viewers and their "true" feelings. Three would make it impossible to compare casual and study group test data.

After each posttest casual viewer had completed the testing, an interviewer asked him, "What will you tell your friends and family about the exhibit?" These comments were recorded on a tape recorder, later transcribed, and rated in terms of positive to negative reaction continuum. Note was made of any comments which suggested a change in the exhibit.

Viewing time, study group. The experimental variations in viewing time were meant to be distributed over the range of viewing times likely for viewers of an exhibit of the size of Vision of Man. Two variations were introduced in viewing time for the study groups. One is referred to as MAX and the other as MIN. A third group, CONTROL, did not view the exhibit at all.

1. MAX group. This group was used in an effort to determine the Maximum Effectiveness Index for the Vision of Man Exhibit. Paid subjects were instructed to look at the exhibit for as long as they wanted and to learn as much as they could. They were subsequently given a comprehensive test.

Their test scores were meant to establish a "ceiling" as to how much learning, attitude change, etc., was possible from the exhibit.

2. MIN group. This group was used in an effort to determine the Minimum Effectiveness Index for the Vision of Man Exhibit. Paid subjects were allowed to look at the exhibit one-half hour and were subsequently tested. This is considerably less time than it takes to read and view all of the information in the exhibit. These data were meant to provide a "floor" and would indicate the amount of change possible if one views the exhibit for only a relatively brief period of time.

3. CONTROL group. These individuals took the same tests as the other subjects, but without viewing the exhibit at all. Their test scores established a base index of knowledge of exhibit topics that may exist in the population without or before viewing the exhibit. By matching the three groups, it could be assumed that the MIN group and the MAX group also had this prior knowledge.

Age/educational level. This was a simultaneous selection of viewer age and/or educational levels; three distinct variations were obtained by selecting study subjects from three populations: high school students, college students, and adults. Casual subjects were measured for age/educational level after the fact, i.e., after they had seen the exhibit.

1. High school students. All subjects who were currently enrolled in high school were placed in this category.

2. College students. All subjects who were enrolled in college or graduate school were placed in this category as long as they did not hold a full-time job. Part-time college students with full-time jobs were placed in the Adults category.

3. Adults. All subjects who were over high school age and not enrolled full time in college were placed in this category. This group was the most heterogeneous with respect to age, education, and occupation.

Since the primary target population of the Vision of Man Exhibit is high school students, emphasis was placed on this category.

Amount of science background. Due to the subject matter of the exhibit, another variable of interest was scientific background. Two categories were established for study groups.

1. Science. All subjects whose background showed an interest and knowledge of science were placed in this category. The criteria used to judge science background was different for each of the three age/education categories.

a. High school. All high school students who had completed at least three high school level science courses were placed in the science category. General science qualified as a science course, but mathematics courses did not.

b. College. All college students who had taken at least one science course per college year were placed in the science category. Nonintroductroy social science courses were counted as science courses, but mathematics courses were not.

c. Adults. All adults employed in the science field were placed in the science category. Remaining adults were categorized according to their educational backgrounds.

   (1) High school graduates. All adults who had not entered college were placed in the science category if they had completed at least three science courses in high school. Again, general science was counted, but mathematics was not.

   (2) College. All adults who had attended college, regardless of level of completion, were placed in the science category if they had completed at least one college science course per year. Again, nonintroductory social science courses were counted, but mathematics courses were not.

2. Nonscience. All subjects who did not qualify for the science category were placed in the nonscience category.

In recruiting study subjects, every effort was made to have an equal balance of science and nonscience subjects in each age/education category. That is, approximately 50 percent of the high school students should be in the science category, approximately 50 percent of the college students in the science category, etc. In assigning the subjects to the three experimental groups -- MIN, MAX, and CONTROL -- the same consideration held. That is, an effort was made to have one-third of the high school science subjects, one-third of the high school nonscience subjects, one-third of the college science subjects in each experimental group. However, as is frequently the case in field work, it was not possible to satisfy all these conditions. As a result, the number actually obtained varied from these ideal requirements. This was especially true of the adult population. Volunteer church and civic organizations were used in most cases (recruitment at a shopping center being the exception). The project staff was literally forced to accept whatever "mix" showed up. This was even more true of the casual group. In their case, background data was used "after the fact" to assign them to an appropriate category; time in the exhibit for this group was a dependent variable rather than an independent variable.

Experimental Design, Exhibit Testing Stage

In accomplishing the objectives set up for the two research stages, each stage required several successive phases. Each phase had a specific purpose and design for analysis.

The Exhibit Testing Stage consisted of two phases thus permitting an initial tryout of the measures developed for empirical effectiveness, followed by a revision and a second tryout. Since the exhibit was on tour, the second tryout was conducted at the exhibit's second site (Chicago). The site, purpose and design specification for these two phases are presented in Tables 1 and 2 in outline and tabular form. The design specification shows the major experimental variations for each phase.

Exhibit phase #1: Development, evaluation, and revision of the empirical effectiveness measures.

Site: Los Angeles, California Museum of Science and Industry
Primary Objectives: To develop the initial versions of the empirical effectiveness measures, collect data on the usefulness of these measures, and revise the measures. See Table 1 for the design specifications.

TABLE 1

Design Specifications; Exhibit Phase #1

			SCIENCE			NONSCIENCE		
			HS	Coll	Adult	HS	Coll	Adult
STUDY GROUPS	CONTROL	M						
		F						
	MIN	M						
		F						
	MAX	M						
		F						
CASUAL VIEWERS	PRE							
	POST							

Exhibit phase #2: Evaluation of the revised empirical effectiveness measures and the generation of tentative hypotheses regarding the relationships among exhibit-related variables.

Site: Chicago, Museum of Science and Industry
Primary Objectives: To collect data on the revised measures for empirical exhibit effectiveness, and to use these data to generate tentative hypotheses regarding the relations among exhibit design variables, exhibit viewer variables, and the empirical effectiveness variables that were investigated during this phase. As can be seen in Table 2, several additional elements were added to the design, particularly in the high school and casual viewer categories.

TABLE 2

Design Specifications; Exhibit Phase #2

			SCIENCE					NONSCIENCE				
			High School			Coll	Adult	High School			Coll	Adult
			HI	MED	LO			HI	MED	LO		
STUDY GROUPS	CONTROL	M										
		F										
	MIN	M										
		F										
	MAX	M										
		F										
CASUAL VIEWERS	PRE											
	POST											

Age/educational level and science background were the only variations measured for casual viewers; socioeconomic level (HI/MED/LO) was investigated only for study viewers who were high school students. This was done by selecting schools from areas considered by those knowledgeable about the city to be composed of families predominantly in high, medium or low socioeconomic levels. It should be noted that socioeconomic level was not identified in the Phase 1 study (Table 1), nor were casual viewers divided by science, nonscience.

Experimental Design, Mock-up Testing Stage

The Mock-up Testing Stage also consisted of two phases. These permitted the validation of the mock-up itself, and the collection of data on experimental variations for several exhibit design variables. Since the exhibit being used in this study was intended primarily for a high school audience and since the Mock-up Testing Stage is concerned primarily with the effects of exhibit design variables (as opposed to audience variables), mock-up testing was, as a justifiable expedient, limited to high school students only. The site, high school(s), purpose, and design specification for each of the two mock-up phases are presented in Tables 3 and 4. The design specifications show the experimental variations for each phase.

Mock-up phase #1: Mock-up validation.

Site: Pittsburgh
Subjects: High school students only
Primary Objectives: To collect data for the validation and/or revision of the mock-up, demonstrating whether or not its effects are comparable to those obtained for the exhibit itself for variations in exhibit subarea, the amount of science background of the viewers, the viewing time, and socioeconomic level. Only study groups viewed the mock-up; there were no casual viewers.

TABLE 3

Design Specifications; Mock-up Phase #1

		HIGH SCHOOL SCIENCE			HIGH SCHOOL NONSCIENCE		
		HI	MED	LO	HI	MED	LO
MOCK-UP (PITTSBURGH)	CONTROL						
	MIN						
	MAX						
EXHIBIT (CHICAGO)	CONTROL						
	MIN						
	MAX						

Mock-up phase #2: Mock-up variation.

Site: Pittsburgh
Subjects: High school students only
Primary Objectives: To collect data on the effects of experimental variations in several exhibit design variables, namely, 1) the amount and readability of text, 2) the amount of visual illustration, and 3) the use of audio versus textual communication. Since this was an exploratory study, only gross and dichotomous variations were employed. Should any of these show significant differences, further investigation could proceed in terms of more continuous variations. Again, only study groups were used in this phase of the study.

TABLE 4

Design Specifications; Mock-up Phase #2

		NO VISUALS	**VISUALS**
TEXTUAL PRESENTATION	**Full Text**		
	Skeleton Text with reduced readibility		
AUDIO PRESENTATION	**Full Text**		
	Skeleton Text with reduced readability		

Empirical Effectiveness Variables - Paper-and-Pencil Measures

Since this study was concerned with the development of research techniques to be used in measuring the effectiveness of an exhibit, several innovative methods of measurement were implemented. However, the basic data re knowledge, attitude and interest obtained for the exhibit casual viewers and paid experimental subjects was collected in a series of seven paper-and-pencil questionnaires. In keeping with the educational orientation of the study, the development of these questionnaires was based on a careful and systematic analysis of exhibit objectives and content. Several phases were involved in this analysis and the subsequent development of the questionnaires. Due to their importance, they are described in considerable detail.

Phase 1. Determining measurable objectives for the exhibit. An exhibit cannot be evaluated in terms of having met its objectives unless these objectives have been well defined. An exhibit objective should ideally state exactly what changes will occur in a viewer as a result of having looked at the exhibit. Evaluation of such behaviorally oriented objectives is relatively easy since what is to be measured has, in fact, already been specified and subsequent changes in behavior can be measured.

In view of this, a preliminary step in the development of the measuring devices was to review the stated objectives for the Vision of Man Exhibit and cast them into specific and behavioral terms. Project staff members met with various individuals responsible for the conception of the exhibit. Discussion centered around the basic aims of the exhibit and what changes in viewer "behavior" could be expected as a result of viewing the exhibit. These discussions helped the project staff determine what objectives the creators of the exhibit thought the "Vision of Man" would accomplish. Their objectives were essentially the same as the five outlined in a mimeographed brochure entitled "Story Line for Exhibit on Federal Science and Engineering for Museum of History and Technology -- Exhibit Objectives," which is reproduced in Appendix A. The exhibit "Fact Sheet" was also a useful source document for this phase of work (Appendix B). The underlying theme of all of this material was the deep commitment the Federal Government has to science and technology; an important secondary theme was the appeal to young people to consider careers in scientific and technical areas and to view the Federal Government in a favorable way as a prospective employer in these areas.

An examination of the five objectives listed in the "Story Line" reveals that they are not stated in terms that readily lend themselves to measurement. For example, the first objective listed stated that the exhibit will "fire the imagination of young people about the impact of science and technology on the world." It is not possible to measure directly either a young person's imagination or how "fired up" it may become by seeing an exhibit. Therefore, this first objective was "translated" to read, "The exhibit will increase knowledge of the basic achievements and impact (technological applications) of science." This increased knowledge would, hopefully, lead to the kind of internal state that may be thought of as "fired up." Table 5 shows the original phrasing of the other four objectives and the attempt to convert these objectives into more behavioral, specific, and measurable statements.

The translated objectives are largely knowledge oriented because this is the only feasible way of measuring such objectives short of long-range follow-up studies of viewer behavior (and even this approach would be subject to gross contamination). In short, one is forced to rely on verbal statements designed to serve as surrogates for the behavior under consideration. It should be noted, however, that along with knowledge items, attitude and interest items were also prepared. While such items present difficulties in interpretation and scoring, they do approach the "feeling" states represented by such terms as "fire the imagination" and "awaken to the excitement" more closely than do items requiring the recall or recognition of factual information.

TABLE 5

Original and Translated Objectives for Vision of Man Exhibit

Original Objectives Listed in Story Line	Translated Objectives Stated in Behavioral Terms
#2 and #3 In developing this world of Federal Science and engineering, we want to show the interrelationship of basic and applied science and technological development and its impact upon man.	#2 To increase knowledge of the interaction between different sciences and between science and technology at both an overall concept level and a more detailed, factual level. #3 To increase knowledge (and regard for) the relationship between the Federal Government and scientific projects.
#4 ... we want to emphasize that scientists and engineers are an essential part of our society.	#4 To increase knowledge about scientists, their projects and the importance of this work.
#5 We want to awaken them (young people) to the exciting current accomplishments in science and engineering and to stimulate them to think seriously of pursuing studies and selecting careers as scientists, engineers, and technicians.	#5 To increase knowledge about careers in science and engineering and the importance of getting young people interested in such careers.

Phase 2. Analyses of exhibit content. The Vision of Man Exhibit was a large, complex display containing information on numerous areas of basic and applied scientific research, most of it Federally supported or accomplished. The amount of written information contained in the exhibit was extensive. A content analysis of all signs and labels was performed before any items were developed for the questionnaires and tests.

All text was copied in the exact format and sequence that it appeared in the exhibit. Individual panels and paragraphs were numbered within each main topic so that the viewing sequence could later be replicated in the mock-up. In addition, the typesize of the lettering was measured and recorded for each panel or paragraph.

The location of all pictures, models and demonstrators were noted for each topic. Audio tape messages were translated into written form. Color slides were taken of all areas of the exhibit, and color motion pictures of areas containing motion. These materials were used as an additional source of information by the staff in developing testing materials.

There were eight primary areas in the exhibit, six of which contained, in turn, topics or subareas. In the content analysis, the subareas were labelled according to topic and were listed under their exhibit areas. A listing of the areas and subareas is contained in Table 6. The physical layout of the Chicago exhibit is shown in Figure 2. (The physical layout of the Los Angeles exhibit, which differed from Chicago in several ways, is shown in Appendix C.)

The total number of words associated with each subarea was counted. The subarea totals were summed to obtain area totals, and the area word totals were summed to obtain the total number of words for the exhibit. This total was checked by a second count of the entire exhibit content.

Once the total number of words for the exhibit had been established, it was possible to compute the percentage each area contributed to the total display. Table 8 (page 30) shows these percentages for each area within the exhibit and for each subarea within an area. These figures were subsequently used to determine the proportions of test items drawn from each subarea and area that would be used to measure gain in knowledge.

A word pool was compiled of all the technical and/or unusual words and phrases used in the exhibit content. This word pool was an essential reference in developing distractors for the multiple-choice questions.

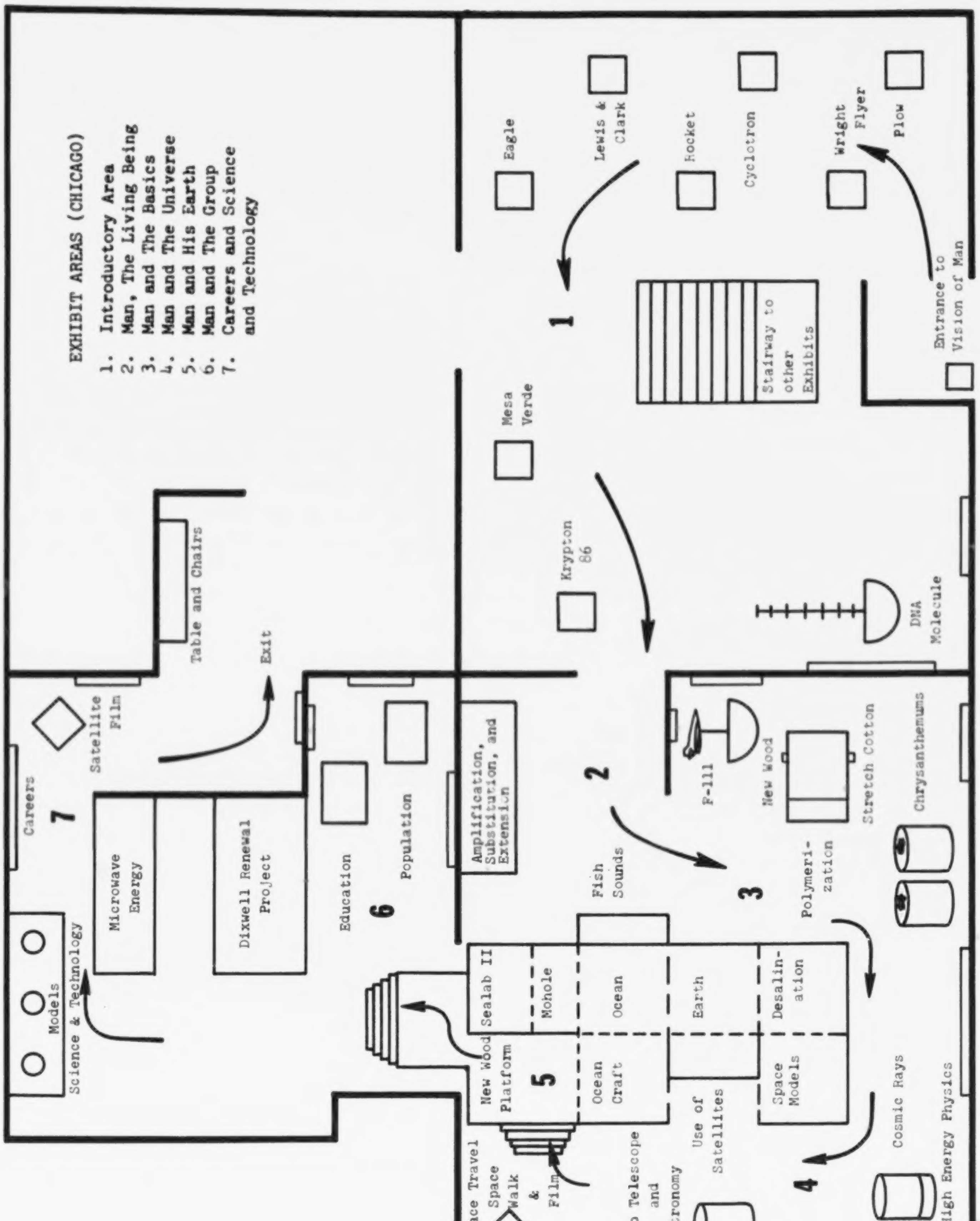


EXHIBIT AREAS (CHICAGO)
1. Introductory Area
2. Man, The Living Being
3. Man and The Basics
4. Man and The Universe
5. Man and His Earth
6. Man and The Group
7. Careers and Science and Technology
Eagle
Lewis & Clark
Rocket
Cyclotron
Wright Flyer
Plow
Entrance to Vision of Man
Stairway to other Exhibits
1
Mesa Verde
Krypton 86
DNA Molecule
Table and Chairs
Exit
Satellite Film
Careers
7
Microwave Energy
Dixwell Renewal Project
Education
6
Population
Amplification, Substitution, and Extension
2
F-111
New Wood
Stretch Cotton
Chrysanthemums
Fish Sounds
3
Polymerization
Models
Science & Technology
Sealab II
Mohole
Ocean
Earth
Desalination
New Wood Platform
5
Ocean Craft
Space Models
Use of Satellites
Cosmic Rays
Space Travel
Space Walk & Film
Radio Telescope and Astronomy
4
High Energy Physics

Figure 2. Physical layout of Chicago exhibit.

TABLE 6

Label and Description of Each Exhibit Area and Subarea

DESCRIPTIVE LABEL	DESCRIPTION OF THE CONTENTS
Introductory Area	This area uses at least one historically significant scientific achievement in each of the subareas to illustrate the effective partnership between science and the Federal Government.
Old Colony Plow	Represents the exhibit area, Man and the Basics. Illustrates the advancements in agricultural science due to Federal Government support.
Wright Flyer	Represents the exhibit area, Man, the Living Being. Illustrates the Federal Government's early interest and support of aeronautical science.
Cyclotron	Represents the exhibit area, Man and the Universe. Illustrates the Federal Government's interest in early experiments in nuclear research.
Rocket	Represents the exhibit area, Man and the Universe. Illustrates the early interest and financing of scientific exploration of space by the Federal Government.
Lewis and Clark	Represents the exhibit area, Man and His Earth. Illustrates Government support of land exploration in the U.S. and throughout the world.

TABLE 6 (Cont'd.)

Pleistocene Eagle	Represents the exhibit area, Man and His Earth. Illustrates Government financing of expeditions to study early life.
Mesa Verde	Represents exhibit area, Man and the Group. Illustrates Government support of anthropological research into the culture of early peoples.
Krypton 86	Represents Man and Measurement. Illustrates Government research involved in precise measurements (no exhibit area).
Man, The Living Being	This area explains recent research into man's effort to protect and prolong life and to understand its origins.
DNA Molecule	Describes research involved in unlocking the secrets of the genetic code.
Amplification	Description of research efforts to "amplify" man's muscle power and illustration of new devices for accomplishing this.
Extension	Description of "skin language" which demonstrates efforts in extending natural sensory capabilities.
Substitution	Description of mechanical devices which can replace or support man's vital organs.
F-111 Aircraft	Description of supersonic aircraft with variable wing.

TABLE 6 (Cont'd.)

Man and His Earth	This area describes man's effort to learn more about the earth and its living creatures.
Fish Sounds	Description of how man is adopting scientific principles by observing porpoises and other sea creatures.
Ocean	Description of recently developed ocean craft and attempts to track the migration of fish as well as mine the ocean floor for valuable mineral resources.
Mohole	Description of Government sponsored project to drill into the earth's surface in order to discover more about early life.
Sealab II	Description of man's recent attempt to live under the sea for extended periods of time.
Man and The Basics	This area describes man's efforts to improve the basics necessary for continued human life.
New Wood	Description of a newly developed irradiated wood which is harder, stronger and more beautiful.
Stretch Cotton	Description of a new material and its unique properties.
Polymerization	Description of the chemical process which "streamlines" molecules into a stronger structure.
Chrysanthemums	Description of controlled changes in plant growth that can be accomplished by use of light and chemicals.

TABLE 6 (Cont'd.)

Phytocrome	Description of the light sensitive pigment in plants which when regulated gives man control over plant growth.
Radiation of Foods and Insects	Description of the beneficial effects of radiating food and harmful insects.
Desalination and Energy Needs	Description of the benefits derived from atomic energy in desalting seawater and increasing available electrical energy.
Man and The Universe	This area describes man's efforts to understand and explore space.
High Energy Physics	Description of research into the atoms.
Cosmic Rays	Description of the cosmic rays, the energy they produce, and information yielded from this research.
Radio Telescope and Radio Astronomy	Description of discoveries about the universe accomplished by the radio telescope and other instruments.
Space Models	Description of spacecraft to be used in exploring the moon and Mars and the engines that will propel them.
Satellites	Description of the basic discoveries in weather and geography made by use of satellites.
Space Travel and Film	Description of the research involved in protecting man during space travel and a film showing an actual space walk.

TABLE 6 (Cont'd.)

Man and The Group	This area describes man's relationships with fellow man in an era of increasing social, economic and technological change
General	Description of the Federal Government's aid in reducing urban problems in housing, transit and slums.
Education	Description of the increasing educational needs of this nation.
Population	Description of world population growth and its demands upon society.
Dixwell Project	Description of urban renewal development in the Chicago area. (Modified for each exhibit site.)
Science and Technology	This area describes the interrelationship between science and technology and shows some technological advancements attained by basic research.
Careers	This area emphasizes the increasing need for scientists, engineers and technicians.

Phase 3. Development of background, interest, and attitude measures.

The background questionnaire: As noted, the design of the study included audience background variables. A questionnaire was drawn up to collect this type of information. The questionnaire was derived in part from one used in the Seattle study (23), but items not directly pertaining to the experimental design of this study were eliminated.

The background questionnaire is included in Appendix D of this report.

The interest index: One of the three main objectives of the Vision of Man Exhibit as identified in Phase 1 work was to increase the "interest" of viewers in science. However, interest, like attitude, is an exceedingly difficult phenomenon to measure. One is generally forced to rely on the ability of human subjects to report accurately their own feelings on the subject. Two factors complicate such studies: 1) people are known to distort their "true" feelings, often in an effort to tell the researcher what the subject thinks the researcher wants to hear (often called the "courtesy bias"), and 2) people are often not really very clear about their own feelings.

In an effort to avoid these difficulties, one could look for substitutes for interest, preferably ones that are easy to observe and measure. In the exhibit situation, one can assume that if a viewer was interested in an area or topic, he would have stopped at that display, examined it, and read at least some of the accompanying text. Therefore, he would have picked up at least some knowledge about the area or topic. The interest test used in this study was based on that assumption. It is included in Appendix E of this report.

The first question was used to eliminate those subjects who were willing to freely admit that nothing in the Vision of Man interested them. Those who answered "yes" to this question were asked to recall from memory the areas that had interested them. The rationale behind this question is as follows:

1. Viewers who had answered "yes" to the first part of the question but who, in fact, had not been interested enough in any of the displays to stop and view them would not be able to remember any areas or topics in the exhibit.
2. Viewers who had stopped and looked at displays, even if they were incapable of learning any of the details of the displays, would be able to remember at least the main topic, or "what the display was about."
3. Viewers who had stopped and looked at a display, and could not remember the topic, must have not been very "interested" in the display.
4. As a corollary to two and three above, if a subject could name or describe in general terms an area or topic of the Vision of Man Exhibit, it must have caught his interest to some extent.

The second question on the interest index was designed to force the subject into making a decision about what topics interest him the most. Unlike the first items, this question could be answered without viewing the exhibit and was used to obtain interest gain or change scores for the study.

The interest index was administered to all experimental groups in the study, both casual and study subjects. Only question 2 was administered to the control groups.

The attitude questionnaire: The items developed for the attitude questionnaire were based entirely on the stated objectives for the exhibit. Each question was designed to measure viewer opinions and attitudes toward one objective. Several items were written for each objective, with the exception of Objective #2 which was measured by only one question, Item #5 in the questionnaire. This was done because it was difficult to formulate other items without repetition of the wording in the original objective. This objective lent itself more readily to knowledge oriented items which were subsequently developed.

Table 7 lists the exhibit objectives and the numbered attitude items written for each objective. The individual items themselves are shown in Appendix F.

TABLE 7

Relationship Between Exhibit Objectives and Attitude Items in the Questionnaire

	Objective	Item Numbers
#1	Awareness of Basic Achievements and Impact (technological applications) of Science	1, 13, 17, 19
#2	Awareness of Interaction between Different Sciences and between Science and Technology	5
#3	Awareness of (and regard for) Relation between Federal Government and Scientific Projects	2, 3, 6, 7, 9, 12, 14, 16
#4	Regard for Scientists and Scientific Projects	8, 10, 11, 15
#5	Interest in Studying Science and in a Scientific Career	4, 18

Several techniques were used in the development of these items. None of the items were specific to the exhibit, although a few of them incorporated some ideas presented in the exhibit. Item #16, which asked which of 20 specialists are employed by the Federal Government, is an example of a question that is easily answered correctly by a person who has viewed the entire exhibit. All 20 specialists had been individually mentioned somewhere in the exhibit. However, this question could also be answered correctly by any person who was aware of the extent of involvement between the Federal Government and scientific projects. Therefore, this objective (#2) was measured by other techniques, one of which was a form of projection. In Item #2, for example, the viewers were asked where they would prefer their son or daughter to work after preparation for a scientific career. A third technique used was asking viewers to choose the type of project they would prefer to be implemented with their tax dollar. This item (Item #10 for Objective #4) lists practical necessities as opposed to basic research projects. Use of their tax money is probably a very real criterion of "worth" for many people.

Phase 4. Development of knowledge measures. Four individual questionnaires were designed to measure knowledge retained by viewers who had seen the exhibit. The rationale supporting the formulation of these questionnaires was that information gained and retained by an individual can be measured within several different levels of learning, ranging from high to low.

The recall testing technique was used to identify a high level of learning. Recall test items are essentially of a "fill-in-the-blank" or "open-end" structure. A correct answer to a recall item indicates successful completion of a three-step process consisting of knowledge selection, integration, and verbalization. That is, after reading the test item, an individual must first select what he thinks are the correct facts from the large repertory "stored" in his memory. The selected facts must then be integrated into a cogent thought. The thought, or idea, is then verbalized. Because of the complexities involved in this process, this type of reiteration from memory is possible only when a high level of learning has been attained; at lower levels, an individual cannot successfully complete one or more of the necessary steps.

The medium and lower levels of learning were measured by two recognition questionnaires. The format of these questionnaires was multiple-choice, consisting of a stem which incorporated the question and four alternative answers, one of which was correct. Recognition items are generally not as difficult to answer correctly as recall items (for a given content area) because the thinking process involved in answering is not as complex. An individual must select knowledge on the basis of four given answers rather than sorting through his entire "fact pool." The process of reading the alternatives may act as a catalyst on his memory to prompt a correct response or aid in eliminating incorrect responses.

Although recall from memory may be considered as an ultimate educational goal, it is unrealistic to expect an exhibit to consistently produce this level of learning. Much of what is learned under these essentially transitory conditions could only be measured by recognition items. However, to measure the full "teaching range" of the exhibit, recall items were considered to be an important part of the test materials.

The open-end concept questionnaire: The items in this questionnaire were designed to measure viewer's ability to recall scientific concepts or principles covered by the exhibit. Conceptual knowledge can best be described as having retained "the big picture" although the individual may not be able to verbalize all the supporting details presented in illustrating the principle or concept. This type of knowledge is the foundation upon which more specific, factual information is built. These items generally asked the viewer to describe the scientific principle involved in a particular exhibit topic. For example, in one item, viewers were asked to describe how porpoises locate objects. The answer "by making sounds," would be considered correct (although more detailed answers would also be considered correct). The open-end concept questions are shown in Appendix G.

The open-end knowledge questionnaire: The items in this questionnaire required a more specific recall of learned information. Item #3 in this test is a correlate to the concept item described above. The viewer was asked to name the technique developed by man for locating hidden objects which is similar to the one used by porpoises. To answer this item correctly, a viewer would have to recall that the echo-location device, SONAR, had something to do with porpoises. This detailed information was incorporated in a subarea exhibit text which explained how man learns from nature, i.e., having observed porpoises, man was able to apply new knowledge in the development of a scientific detecting device.

In most of the individual content texts, there was a typical sequence for presenting information. Most often, the scientific principle or concept was first explained. Specific facts which supported the principle were then presented. Therefore, the detailed facts required to correctly answer most open-end knowledge items appeared in the latter half of an exhibit text, which meant that a viewer had to consider the concept text interesting enough to continue reading. The open-end knowledge items were developed to simultaneously examine viewer retention of specific facts and the influence of exhibit content sequencing on learning. The open-end knowledge questionnaire is shown in Appendix H.

The multiple-choice questionnaire: Since a large percentage of the knowledge questions developed for this study were incorporated into the multiple-choice questionnaire, several preliminary procedures were followed prior to item construction. The first step was determining the exhibit area and subarea percentages by using the word count data. The project staff then converted the area word percentage directly into the number of items that should be written for that area. For example, the exhibit area, "Man and His Earth," contained a total of 1757 words which accounted for 18 percent of the entire exhibit. Therefore, assuming a 100-item test, it was determined that 18 out of 100 items would be constructed for the topics contained in that area. The actual number was 17. The number of items developed for each subarea was also roughly proportionate to the obtained subarea percentage. For example, the subarea, "Ocean" represented 27 percent of the "Man and His Earth" area. Since this was the largest subarea percentage, four multiple-choice questions were developed for "Ocean." Table 8 shows the number of multiple-choice items written for each exhibit area and subarea. (The rationale supporting the exhibit-only items is explained in the next section of this report.)

A technical word pool was developed for use in devising the three incorrect alternatives for each question. Each item, which incorporated four alternatives taken directly from the exhibit content, forced the viewer into weighing each answer against the stem rather than choosing the alternative which was easily recognized as having been seen in the exhibit. This method of item construction helped to assure the internal consistency of the questionnaire. In the final multiple-choice tests which were used in the Chicago field tryouts, 72 percent of the alternatives were developed directly from the exhibit technical word pool. An example of the development of a question directly from the exhibit text and the use of exhibit related alternatives is multiple-choice item #4. The text reads, "The multi-purpose F-111 was built around the concept of the variable wing." The question reads:

> In designing the F-111, engineers primarily made use of knowledge gained from:
>
> a. the development of the ion engine
> b. Goddard's paper on reaching extreme altitudes
> c. experiments with the variable wing
> d. the field of rocket astronomy

The three wrong alternatives were all taken from the exhibit and were in some way connected with aeronautics, but in no way could be considered the correct choice for this item.

Prior to the first field tryout, 100 multiple-choice items were written using the guidelines outlined above. Since many areas of the exhibit were quite technical and required extensive knowledge for complete understanding, the staff submitted certain items and copies of the exhibit text to consultants who were experts in the various technical areas. These consultants, having reviewed the items and text, made many suggestions for item revisions. Such revisions assured that the alternative designated as correct was in fact correct, and the other three alternatives were in no way correct.

An additional analysis was performed on the original items before the field tryout. A check was made on the difficulty level of each item. Three members of the staff individually examined each

TABLE 8

Chicago Word Count, Area Percentages, and Number of Items for Multiple-Choice and Exhibit-Only

	# Words	%	# MC*	# ExO**
Introduction Panel	68	27	0	1
Exit Panels	183	73	0	1
(3% of Exhibit)	251	100	0	2
MAN AND HIS EARTH				
Eagle	84	5	1	1
Lewis and Clark	134	8	1	0
Fish Sounds	368	21	3	0
Ocean	481	27	4	0
Mohole	324	18	3	1
Sealab II	166	9	2	0
Mining Ocean Floor	200	11	1	0
TOTALS	1757	99	15	2
(18% of Exhibit)				
MAN, THE LIVING BEING				
Wright Flyer	114	8	1	0
DNA Molecule	470	34	4	3
Amplification	151	11	1	2
Extension	115	8	0	1
Substitution	142	10	1	0
F-111	373	27	3	0
TOTALS	1365	98	10	6
(14% of Exhibit)				

TOTAL WORDS = 9809

	# Words	%	# MC	# ExO
MAN AND THE BASICS				
Old Colony Plow	120	5	1	0
New Wood	183	8	1	0
Stretch Cotton	132	6	1	0
Polymerization	106	5	1	1
Phytochrome	124	5	2	0
Chrysanthemums	235	10	0	1
Mutations	328	14	2	0
Radiation -- Insects and Food	277	12	5	1
Desalination	268	12	1	0
Energy Needs and Sources (Microwave Energy)	428	19	7	1
Wood Flooring	96	4	0	1
TOTALS	2297	100	21	5
(23% of Exhibit)				
MAN AND THE UNIVERSE				
Rocket	76	3	1	0
Cyclotron	94	3	1	0
Atom	591	22	4	1
Cosmic Rays	148	5	2	1
Theories	61	2	0	1
Radio Telescope and Astronomy	585	21	4	0
Space Wrench	219	8	1	1
Space Travel	260	10	1	0
Nerva	58	2	1	0
Molab	118	5	1	0
Gulliver	42	2	1	0
Mariner IV	31	1	1	0
Chemical and Ion Engines	153	6	1	0
Satellites	300	11	5	0
TOTALS	2736	100	24	4
(28% of Exhibit)				

*Multiple-Choice

**Exhibit-Only

TABLE 8 (Cont'd.)

MAN AND THE GROUP	# Words	%	# MC	# ExO
Mesa Verde	105	15	1	0
General	253	35	1	0
Education	156	22	1	1
Population	200	28	0	0
Urban Renewal	0	0	0	0
TOTALS	714	100	3	1
(7% of the Exhibit)				
MAN AND MEASUREMENT				
Krypton 86 Gas	203	2	2	0
(2% of Exhibit)				
CAREERS AND SCIENCE AND TECHNOLOGY				
Careers	109	22	1	0
Achievements of Science and Tech.	377	78	5	0
TOTALS	486	100	6	0
(5% of Exhibit)				

item and rated it on a difficulty scale ranging from one to three; a three rating equalled "most difficult." The criteria used for this rating were based on the location of the correct answer in the text, how much it was emphasized in the text, and the typesize in which it was printed. The ratings were summarized for the items in each subarea. Most items were rated "one" or "two." A perfect distribution for an area appeared when most items clustered at "two" level with a few at "one" and "three." Each item and difficulty level was discussed by the staff. Subsequent revisions of items were based on the overall difficulty of each topic and subarea. Alternatives which were unnecessarily easy or difficult were changed.

The multiple-choice items were revised again following data compiled from the first field tryouts. All tests were scored for the study and control subjects. An item analysis was performed for all multiple-choice items, summarizing the total number of people in each experimental group that chose each alternative. Any item for which an incorrect alternative was selected more frequently than the correct alternative by the total study group was automatically revised by the staff. A chi square test was performed on all other items. This statistic was used to determine whether subjects who rejected the correct answer chose equally among the three incorrect distractors, i.e., did each incorrect alternative "pull its weight?" When a significant chi square was obtained for an item, it indicated that one of the incorrect distractors was more attractive to the respondents than the other two. All such items were reviewed by the staff. In some cases, extensive revisions were made in an attempt to equalize the "pulling power" among the incorrect distractors. In summary, the use of an item analysis and the chi square technique were of help in isolating poorly constructed items and pinpointing the specific difficulty.

The final distribution of multiple-choice items was not "ideal," i.e., the percentage of items for each subarea did not always correspond exactly to the word count, nor was there always the correct distribution of difficult and easy items. However, the approach described here represents an effort to apply systematic procedures to item construction. Without such an effort, little can be said about the real effect of the exhibit on the viewer, particularly when questions relating to specific subareas are raised or when the total impact of the exhibit is of concern. The multiple-choice questions are shown in Appendix I.

The exhibit-only questionnaire: It was intended that exhibit-only items would be the least contaminated by "prior knowledge" and would thus permit analysis of learning without requiring control group or pretest correction. Knowledge to answer these items could only be obtained by recalling a specific item from the exhibit itself (e.g., what model was used to illustrate a certain principle), and could not be answered on the basis of general knowledge of the subject matter. Exhibit-only items were also used to fill in gaps in the multiple-choice test, that is, areas that were not adequately represented in proportion to their word count. For this reason, the

20 items developed for this test were included as part of the 100-item multiple-choice test. The same statistical techniques were used to revise these questions as those explained in the previous section. The exhibit-only questionnaire is shown in Appendix J.

Empirical Effectiveness Variables - Nonpaper-and-Pencil Measures

As noted earlier, in addition to the paper-and-pencil tests, a number of other techniques were used to measure exhibit effectiveness. Since the aim of this study was exploratory, it was considered important to try as many approaches as could be implemented within the time and budget limitations of the project. These techniques will now be described.

Video tape analysis. There are many design elements in a complex exhibit, each of which may make a contribution to its overall "attracting" and "holding" power. One dimension along which such elements could be classified is static and dynamic. A static design element is one which stands in a fixed position in a display; immobile models, signs, and photographs are examples of such elements. Dynamic elements are moving or changing; models and demonstrators with moving parts, audiotapes, films, and special lighting and sound effects are examples of these elements.

In another section of this report these elements are discussed in the context of exhibit design variables. That is, an internal measure of attracting power is described, based on the existence and number of static and dynamic models in each exhibit subarea. The dependent variable is the number of visitors who stop. In short, the question asked is, is there a correlation between the number of static and dynamic models and the attracting power of the exhibit.

The present discussion looks at attracting power from a different direction. It considers the development of an empirical measure based on direct observation of visitors as they observe a display, noting what, in fact, they look at as they scan the various elements.

A number of studies have been done in a related field. Hess and Polt (13) noted that the size of the pupil reflects the interest level of a person observing an object, i.e., the larger the pupil size, the higher the interest level. Hess (15), in conjunction with Marplan Perception Laboratory, a market research organization, developed a portable eye camera which is attached to a viewer's head as he views the various objects. The eye camera enables the researcher to determine exactly what the viewer is looking at and his interest as measured by pupilary reaction.

Consideration was given to adopting the portable eye camera to measure the "attracting" power of each design element of the exhibit. However, this approach proved impractical on several counts. First of all, the group to be investigated was the casual viewer group.

If an eye camera were attached to their heads, they would no longer be "casual" viewers, but experimental viewers. Secondly, to minutely measure each design element of each display would perhaps be ideal in determining an absolute value for each element, but such values would be of little help in evaluating the relative attracting power of each element as part of the integral display. Thirdly, overall cost of the technique made its use prohibitive.

Although the staff rejected the use of the eye camera as a technique for measuring the "attracting" power of design elements, the detailed evaluation of the technique did serve to draw attention to criteria that any measuring device should have:

1. The "casual" viewer must be unaware that his reactions are being recorded.
2. The whole display area should be exposed in order that all design elements have an equal chance of attracting the viewer.
3. The cost and the time for data collection and subsequent analysis should be kept at a minimum.
4. The results of the measurement should be permanently recorded for reference during analysis.

A video tape recording system was felt to come the closest to meeting these requirements. The general plan was to record on video tape casual viewers actually looking at the exhibit, but without their knowledge. The methodological question was, would it be possible to determine from such tapes what portion of the display they were viewing in a reliable fashion. If so, the technique would provide a stable basis for analyzing viewer behavior in the "real world" with respect to the relative attracting and holding power of the various display elements contained within the exhibit. Because of the potential usefulness of such a technique, the procedures followed in this phase of work are described in some detail.

During the Chicago field tryouts, an Ampex video tape system was installed to record casual viewers' reactions to four distinct exhibit display areas. A television camera was hidden within each display area; the television monitoring and tape system was concealed behind a screened-in area away from the display. From this location, a staff member could observe casual viewers on the television monitor. The video tape machine itself was operated only when it appeared that a casual viewer was approaching the display area.

To develop the appropriate means to analyze the data, a location reference tape was made of each exhibit. While one person operated the tape device, another staff member looked at and pointed directly to each model and sign in the exhibit area in a predetermined order. Using a microphone and the audio channel of the tape, it was possible to thus number and identify each design element consecutively, recording the element number on the audio track of the video tape.

After the reference tapes had been made, approximately one and one-half days were spent taping casual viewers at each display area. The amount of time depended upon the number of casual viewers who happened to view each display. There were often long periods of time when no taping could be done because of the scarcity of viewers.

There were a number of steps necessary before actual analysis of the video tapes could be accomplished. First, the four reference tapes were given to a graphic arts specialist. Using the Ampex tape system and the television monitor, the specialist photographed the staff viewer looking at each numbered display element. Each photograph showed the position of a "viewer's" head and face as he observed any one display element. These photographs for each exhibit display were then arranged sequentially around the periphery of a photograph of the entire exhibit area showing all the various display elements. Lines were drawn from the numbered position of the viewer to the exact display element in the center picture. Therefore, by following the line from the photograph of the "viewer's" head position to the center picture, one could (hopefully) identify the display element that was being viewed by a "real" casual viewer.

The second step in the tape analysis was to assign consecutive subject numbers to the taped casual viewers so that they could be positively identified in the subsequent analysis. While watching the television monitor, a staff member recorded on the audio track subject numbers for viewers who looked at any one part of the display. For purposes of this analysis, a person who was just passing through the display area was not considered a subject and was not assigned a number. (As noted, an effort was made to tape only "real viewers." However, it was frequently the case that a large group would be around the display and it was necessary to identify and distinguish between those who were attracted to the exhibit, even if for only a very short time, and those who were just "passing by." Also, several young viewers spotted the small video camera hidden within the display and immediately began putting on "acts" for the benefit of their unseen audience. While they were definitely attracted to the exhibit, it seemed to be for the wrong reason, and they were not included in the analysis.)

Two other preliminary steps were necessary before actual analysis. The outlines and numbers assigned to display element positions contained within the exhibit were drawn on the television monitor screen with a grease pencil. This was an additional aid for determining where a subject was looking. These markings naturally had to be changed for each tape. A third piece of equipment, an Esterline-Angus Event Recorder, was used to record the frequency with which each display element was viewed by a given subject. The event recorder consists of a moving graph with twenty pens producing vertical lines. A keyboard of twenty corresponding buttons was used to activate the pens. When button #1 is pressed, a small tic mark is made on line #1 of the graph. In this analysis, such a mark on vertical line #1 meant that a casual viewer had looked at display element number one in the exhibit.

To determine the reliability of this approach, each exhibit tape was rated independently by three judges. Several operations were involved in analyzing each tape. A judge turned on the video tape recorder and television monitor, wrote the subject's number on the Esterline-Angus graph paper, and then started both the recorder and the graph. He simultaneously watched the television monitor and pressed the appropriate keys for each subject. If there was any doubt as to exactly where the viewer was looking, the judge could stop the recorder and refer to the position locator pictures for a more exact judgment. He could also reverse the tape as often as required. With some practice, a judge was able to observe both the television monitor and the position locator pictures while operating the graph. Each judge performed all of the above operations in rating all subjects for the tapes analyzed. After a judge had completed his ratings for a display, the graph paper was removed from the Esterline-Angus. A check was made to assure that all three judges rated the same numbered subjects. (See Figure 3 for a view of the judging area. Several of the items described above can be seen, including the tape machine, the monitor, the Esterline-Angus, and the location aid.)

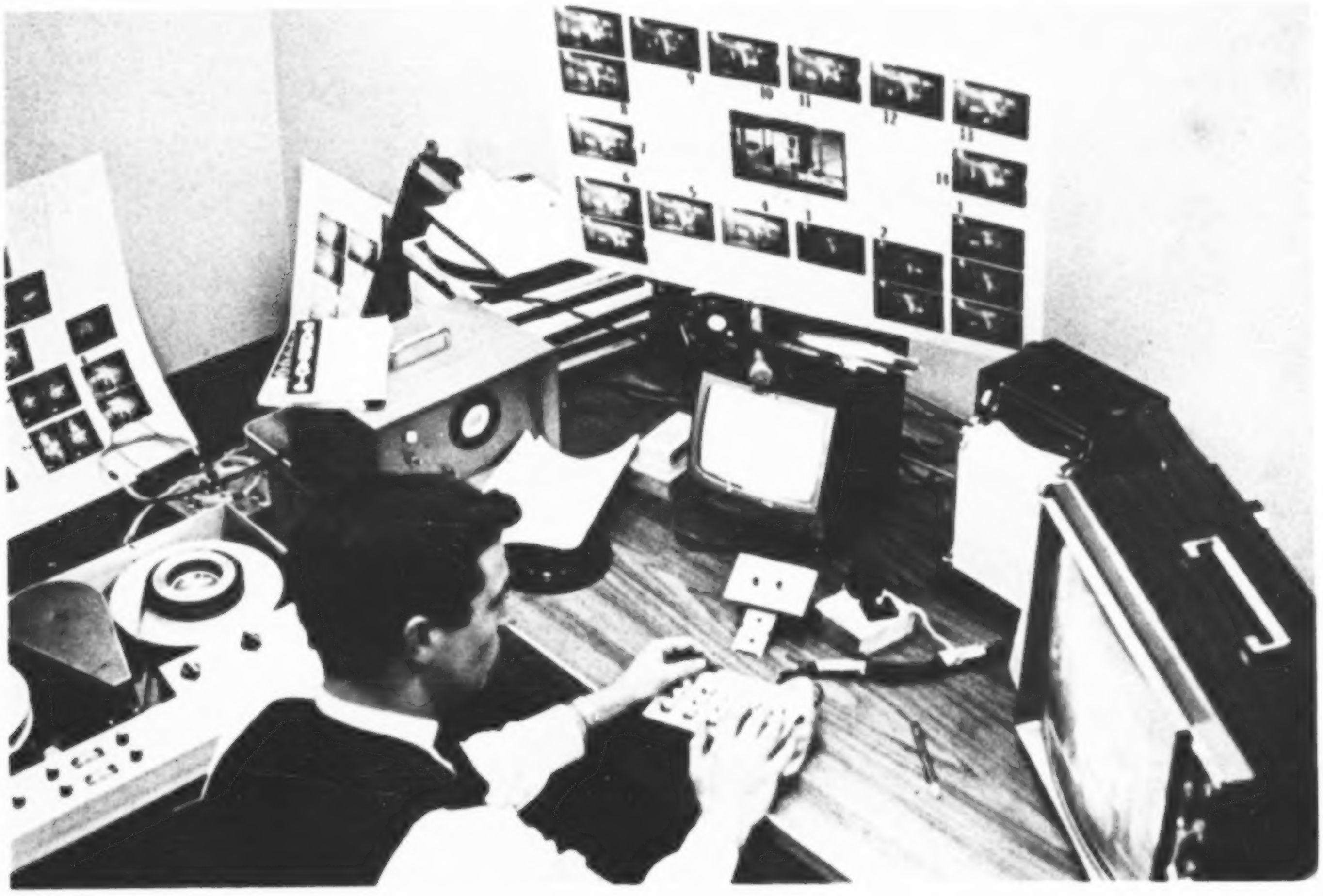


Figure 3. View of judging area showing judges, Ampex Video Tape System, "Fish Sounds" placard, television monitor, Esterline-Angus Event Recorder, and an additional television monitor.

Only two of the tapes, Fish Sounds (N=35 viewers) and Desalination (N=36 viewers), were rated by the three judges. After the ratings were completed for these areas, a preliminary analysis was performed on the data. The results were significant enough to conclude that the additional time involved in rating the other two tapes (approximately four days) would not greatly enhance the evaluation of this technique. Also, Space Models and Mohole, the two tapes that were not rated, were of poorer quality, i.e., they were located toward the end of the exhibit area and were therefore not as well "populated" with casual viewers. Also, neither of these displays contained any dynamic display elements. In the Space Models exhibit, it was more difficult to estimate where the viewer was looking because of the density of the design elements in a relatively small area. Because of an obstructing design element, it was necessary to place the television camera to the left of center in the Mohole display. Such placement of the camera created deceptive parallax angles and made it difficult to judge where the viewers were looking. Consequently, it was not possible to accurately draw the outlines of the display elements on the television monitor and the resultant ratings would have been unreliable.

Readability analysis. As has been noted, several key exhibit variables were considered as having an impact on learning and attitude and interest change on the part of the viewer. One of these variables was the difficulty level of the textual material. It would seem obvious that the easier the text is to read, the more information the viewer would get out of it, and vice versa. Reading level should thus correlate with scores on the questionnaire, particularly knowledge items. In order to test this hypothesis it was necessary to establish the grade level of the textual content for each of the subareas and areas contained in the exhibit. The established grade levels are equivalent to reading grade levels in years of schooling, i.e., a high school sophomore should be reading at a 10.0 grade level.

The technique used to determine these grade levels was the 1948 Revised Flesch Readability Formula. This formula takes into account average word length and average sentence length to arrive at a figure indicating the difficulty level of the text. There are actually two formulae, one giving a grade level and the other giving a figure called the Flesch Count. The higher the grade level, the more difficult the text; the lower the Flesch Count, the more difficult the text. Grade levels are shown in Table 9 for each area and subarea. Grade levels in the Vision of Man Exhibit ranged from 6.8 to 14.8.

The measurement of typesize. While written content of the exhibit was being recorded, an analysis was also made of the typesize of each label, sentence, paragraph, etc. In all, 14 different typesizes were identified, ranging from 1/8 inch to 2-3/4 inches. (The average typesize for each area and subarea is shown in Table 9.) It was felt that this variable may contribute to learning in a consistent pattern, i.e., the larger the type, the more learned. However, a preliminary analysis indicated an inverse correlation between typesize

TABLE 9

Number of Words, Reading Grade Levels, and Average Typesize in Exhibit Areas and Subareas

Description	# Words	Avg. Grade Level to 10ths	Avg. type-size 100ths of an inch
Area-Introduction	930	10.9	.29
Eagle	84	9.4	.30
Lewis & Clark	134	11.1	.28
Plow	120	12.7	.27
Mesa Verde	105	11.1	.29
Wright Flyer	114	9.3	.29
Rocket	76	11.9	.30
Cyclotron	94	11.9	.29
Krypton 86	203	10.3	.27
Area-Man, the Living Being	1,251	10.3	.32
DNA	470	10.9	.40
Amplification, Extension, and Substitution	408	10.3	.24
F-111	373	9.5	.32
Area-Man & His Earth	1,521	9.8	.49
Fish Sounds, etc.	368	9.5	.24
Ocean (Subs & Fish Migration)	481	10.7	.59
Mohole, Sealab, & Mining	672	9.3	.55
Area-Man & the Basics	2,177	10.2	.32
Molecular Δ & New Wood	517	10.3	.37
Agriculture	964	9.9	.26
Desalination & Energy Needs	696	10.8	.35
Area-Man & the Universe	2,566	9.8	.31
High Energy Physics	591	10.6	.29
Cosmic Rays	148	8.9	.13
Atmosphere	646	10.9	.34
Space Models	662	9.0	.33
Space Wrench	219	6.8	.37
Scientific Satellites	300	10.9	.31
Space Walk (Film)	—	—	—
Satellite Film (Score)		—	—
Area-Man & the Group	679	9.3	—
General	253	9.0	.38
Education	156	10.1	.25
Population-Chicago	200	8.7	—
Urban Development-Chicago	70	—	—
Area-Careers & Science & Technology	486	11.4	.44
Careers	109	14.8	.42
Science & Technology	377	10.4	.44
Area-Intro & Exit	251	11.1	.73

and grade level; that is, the smaller the typesize, the higher the grade level. This correlation is not as adventitious as it may at first seem. Larger typesizes are used for main headings and labels, which are usually short and to the point. Detailed explanations are generally printed in smaller type underneath. The more technical the information to be included, the smaller the typesize generally employed. Unfortunately, however, this relationship meant that both factors would be operating on knowledge gained and a controlled study would have been required to separate out typesize and readability factors in order to test their influence individually. While such a study would most certainly be a worthwhile endeavor, it was not feasible within the scope of this project. Therefore, the treatment of typesize as a separate factor was dropped. It should be remembered that it is an intrinsic part of the readability analysis.

Time data. Casual viewer time data was collected for each subarea of the exhibit. Field interviewers stood at each subarea until 60 casual viewers had looked at the display. Time spent at the subarea was kept in seconds; high school, college, and adult time data were recorded separately. These samples of 60 casual viewers was different from the posttest group who had been tested.

In Chicago, data was also collected on the number of casual viewers who stopped at each subarea. Sixty casual viewers were individually followed through the exhibit and the order in which they stopped at the subareas was recorded. The subjects in this group were also different from the posttest group.

These time data provide an index of both "attracting power" and "holding power" for each area and subarea of the exhibit. These measures were subsequently used as dependent variables in several analyses having to do with casual viewer behavior.

Mock-up Testing

The design of both mock-up phases of this study have already been described (pages 14 through 15). A fuller discussion of the rationale for this work, and the procedures followed in carrying it out, is given here.

Designing and building scientific and technical exhibits on as large a scale as the Vision of Man is time consuming and costly. Once such an exhibit has been built and is ready for public "unveiling," any errors in design, content, and/or sequencing become, by and large, a permanent fixture of the exhibit.

It was hypothesized that if a small-scale simulation or mock-up of a planned exhibit were built and tested for effectiveness prior to exhibit construction, errors could be located and modifications made before being committed into a permanent structure.

Designers would then be assured that the permanent exhibit, although still costly, was, in fact, accomplishing its objectives.

[This approach is again in keeping with the modern educational orientation of this study. Programmed instructional materials, for example, are pretested and revised as a matter of course in an effort to achieve their intended level of performance. Gropper (7) has applied the technique to educational television and shown that pretesting definitely improves the final product by avoiding serious mistakes in content, sequencing or approach.]

The hypothesis stated above depends upon an important assumption: To what extent does a simplified, low cost model of an exhibit achieve results comparable to those achieved by the real exhibit? If the interest, attitude, amount learned, etc. are comparable, then such a mock-up could be used as hypothesized, i.e., as a method of pretesting the final product. If the mock-up does not relate in a significant way to the actual exhibit, then this approach would be unsound, and the hypothesis must be rejected. Furthermore, if the mock-up was related to the exhibit, it could be used as an experimental medium for studying exhibit effectiveness. The ease of manipulating the mock-up in comparison to real exhibits would result in considerable savings in conducting such research. The potential value of this approach warrants the fairly detailed description of its implementation in the pages that follow.

A mock-up of the entire Vision of Man Exhibit was created from the recorded text content and Kodacolor photographs collected at the exhibit sites. Thirty-one individual panels were designed to represent all subareas of the actual exhibit. Replications of each exhibit subarea were produced by mounting all pertinent photographs of subarea models and pictures along with corresponding texts onto large pieces of artboard. Care was taken to assure that the juxtaposition of all photographs and text passages were essentially the same as in the actual exhibit. Figure 4 shows the actual exhibit subarea, DNA, and a picture of the DNA mock-up panel. Figure 5 shows the mock-up of one of the introductory displays.

Two limitations were inherent in the mock-up design. The mock-up was, of course, two dimensional. Secondly, it was not possible (nor desirable from a cost standpoint) to replicate all moving models and exhibit features in which viewers could participate (e.g., pick up a phone to hear a sound track, press a button to make a model operate, etc.). In fact, the inclusion of such sophisticated devices would invalidate the notion that a simplified, two dimensional mock-up can simulate the actual exhibit.

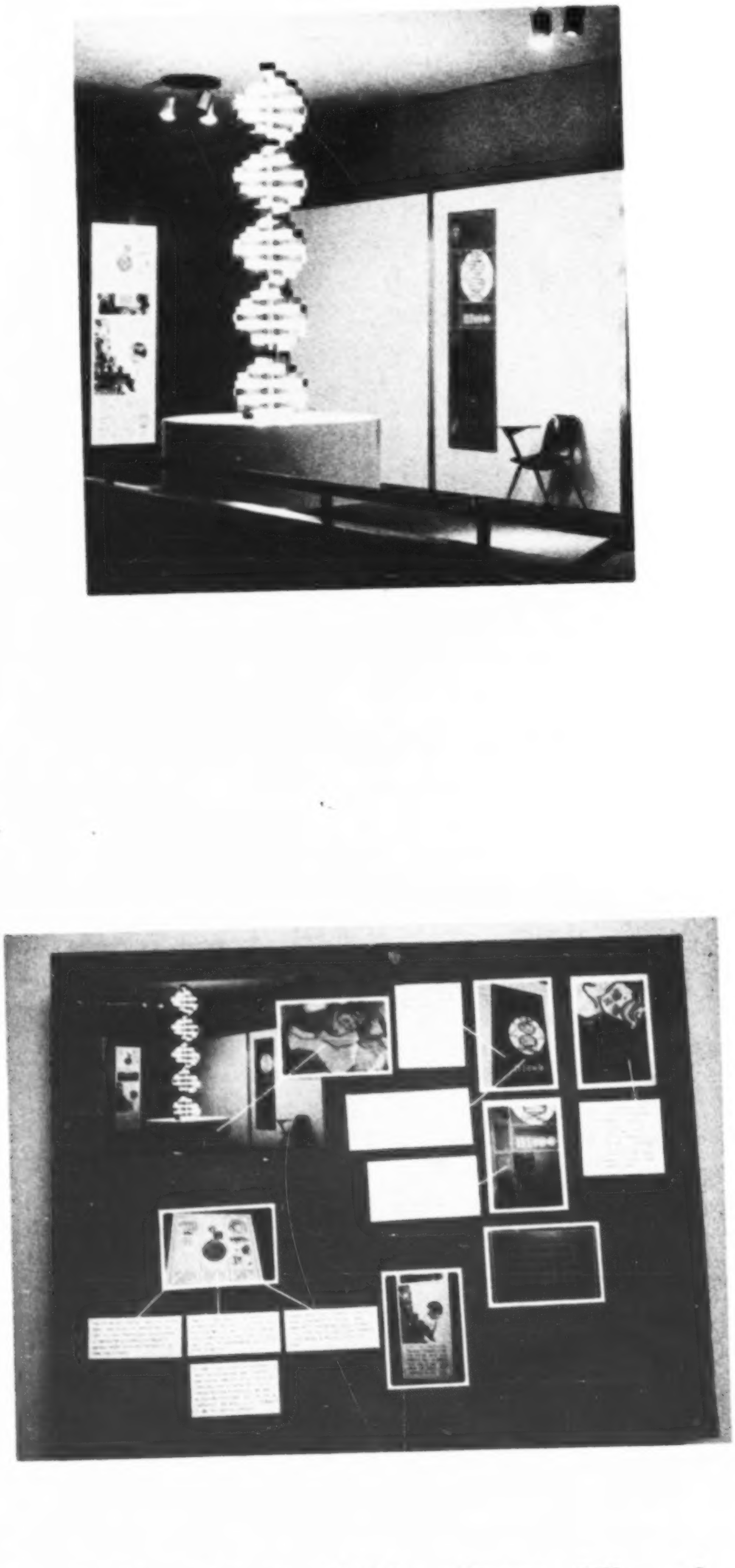


Figure 4. Photographs of exhibit subarea, DNA, and corresponding DNA mock-up panel.

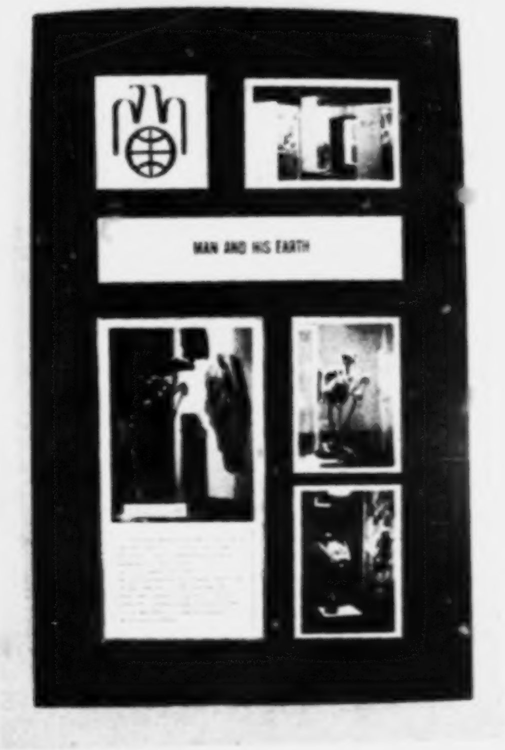


Figure 5. Mock-up panel of Introductory Display, Man and His Earth.

It may also be noted that where a color photograph distinctly showed the text of a sign or label, it was used as the text in the mock-up. Where this was not possible, the text was typed and made a part of the mock-up. These typed passages were formatted the same as in the exhibit, but typesize was not varied. (See Figure 4 for an example of how this was accomplished for the DNA mock-up panel.)

The building of a mock-up for this exhibit established that such a course of action is feasible in terms of time and money. It should be noted, however, that this mock-up was created after the fact; that is, the "building ingredients" were extracted directly from the existing exhibit. Therefore, at this point, no statement can be made regarding the feasibility of mock-up construction prior to exhibit construction. This would be a worthwhile area of investigation in a future research study.

Mock-up testing consisted of two distinct phases: validation and variation. During the actual exhibit testing stage, the age and educational level of the viewer audience was varied in order to explore the effects of these variables. However, since the exhibit was intended primarily for high school students, mock-up testing was limited to only high school subjects. The site used in both phases of the mock-up testing was the American Institutes for Research building, located in Pittsburgh, Pennsylvania. All subjects were from high schools located within the Pittsburgh area.

Phase 1. Mock-up validation testing. The amount of mock-up exposure time was controlled with the experimental groupings of MIN, MAX, and CONTROL. Subjects were selected on the basis of the science/nonscience dichotomy with 30 students in the science category and 32 in the nonscience group.

The 62 subjects were selected from three area high schools, i.e., approximately 20 students from each school. The schools were selected so that one socioeconomic level was represented by each school. The students tested from School A were from a high socioeconomic background; School B from a medium level; and School C from a lower level.

The overall layout of the mock-up subarea panels followed the basic sequencing of topics in the actual exhibit. Figure 6 shows a diagram of the mock-up layout. The introductory panels of the mock-up were hung together in one room; this was comparable to the exhibit introductory cases. All other mock-up panels were hung consecutively along three main corridor walls. Study subjects were lead into the introductory room by a route which did not go through the mock-up area. This procedure, which ensured the integrity of attitude and knowledge gains within the framework of imposed time limits, had also been followed during exhibit testing tryouts with study groups. Figure 7 shows study subjects viewing the mock-up.

Prior to the actual testing, each high school had been asked to select participating students on the basis of science or nonscience background, limiting the number of students to ten in each category. The three high schools were tested on three different Saturdays.

The testing procedure was as follows: Upon arrival, a staff member introduced three testing monitors and then gave a general description of the project to the students. The participating high school usually provided a listing of students' names under the science or nonscience categories. Within each category, students were randomly assigned to the MIN, MAX, and CONTROL experimental groupings. The MIN and MAX groups usually consisted of eight subjects each; four science and four nonscience. The CONTROL group was limited to four subjects each; two in each category. All study subjects were given group identification badges indicating MIN or MAX. Each group was assigned a staff monitor who supervised the group throughout the whole viewing and testing procedure. The MIN group was the first to view the mock-up. After they had viewed the introductory area, the MAX group was allowed to start their viewing. When both groups had left the conference room, CONTROL subjects were administered the

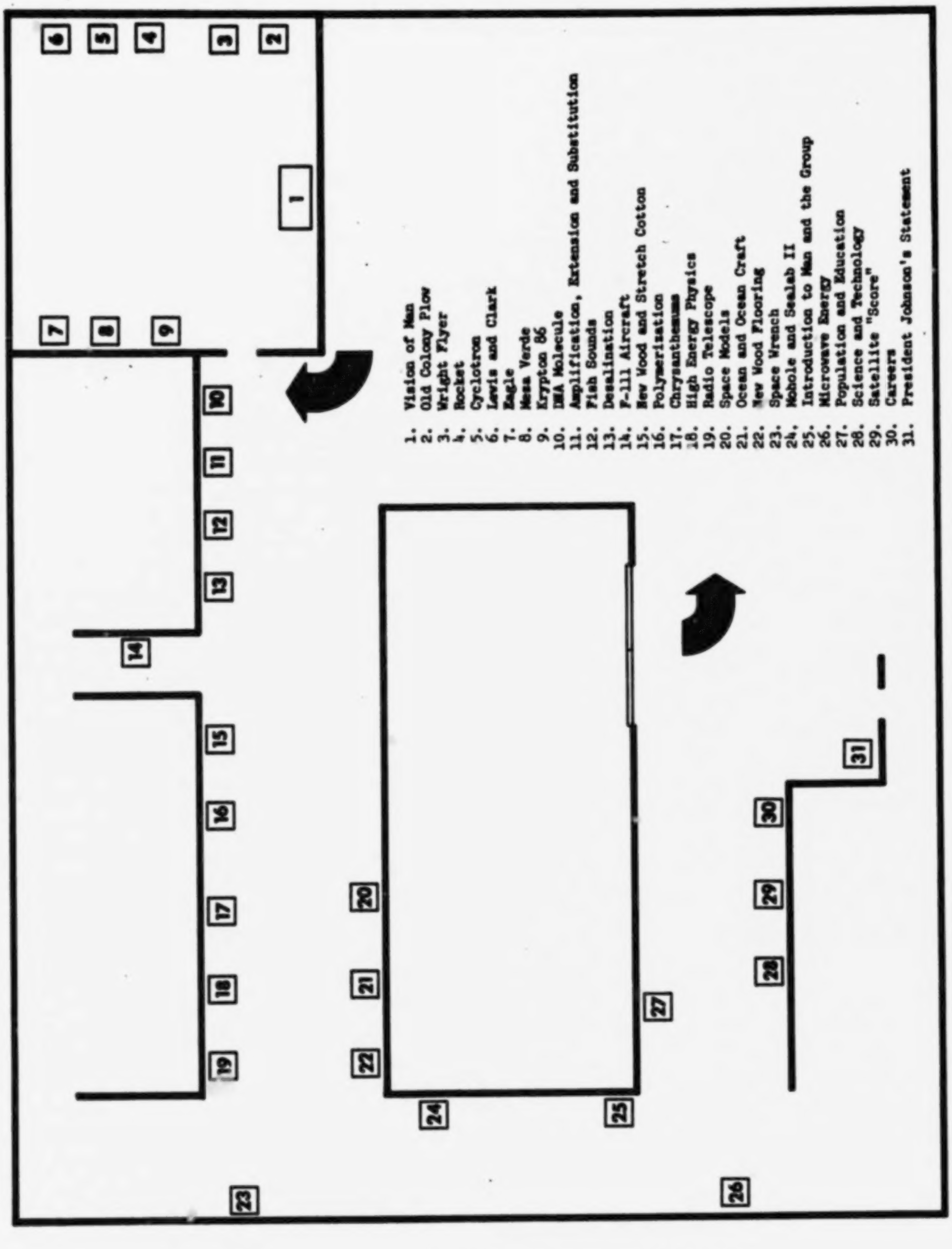
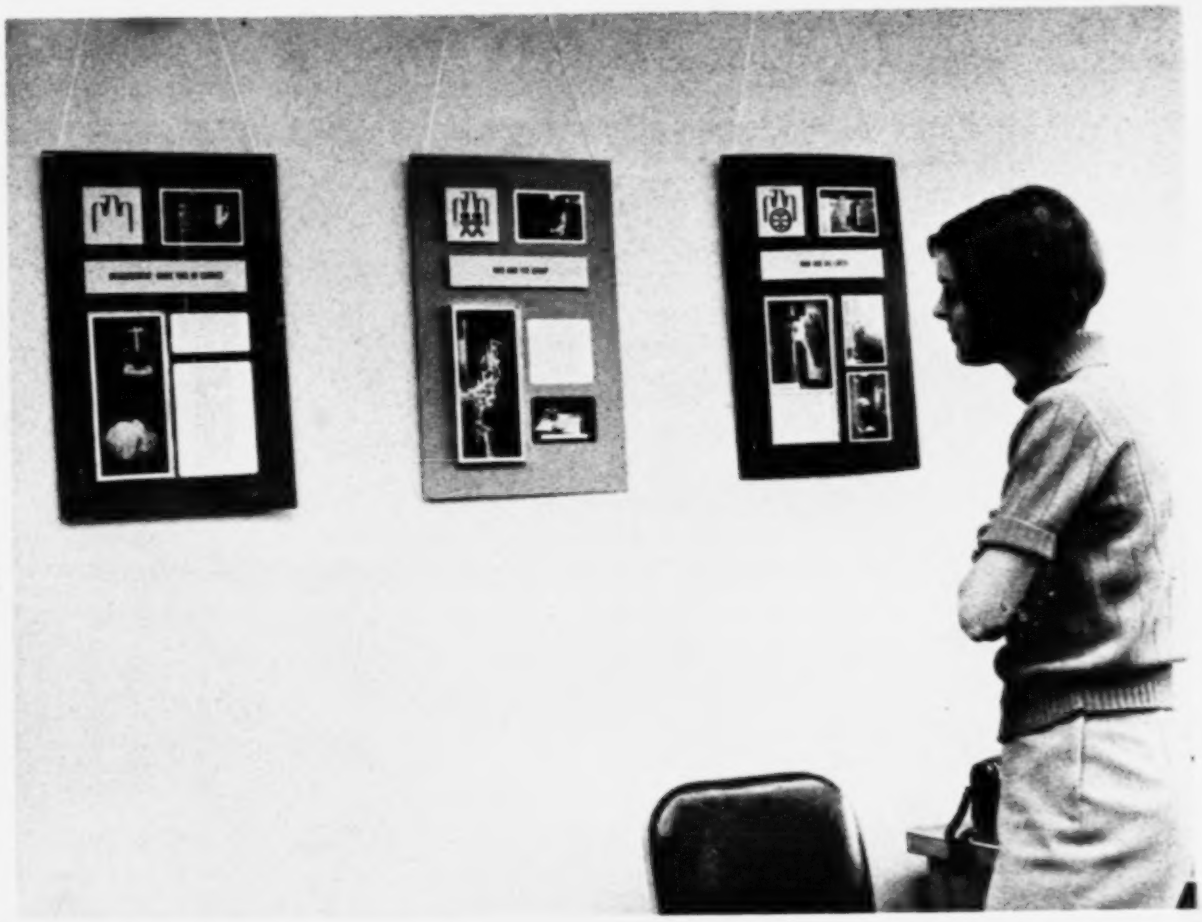

1. Vision of Man
2. Old Colony Plow
3. Wright Flyer
4. Rocket
5. Cyclotron
6. Lewis and Clark
7. Eagle
8. Mesa Verde
9. Krypton 86
10. DNA Molecule
11. Amplification, Extension and Substitution
12. Fish Sounds
13. Desalination
14. F-111 Aircraft
15. New Wood and Stretch Cotton
16. Polymerization
17. Chrysanthemums
18. High Energy Physics
19. Radio Telescope
20. Space Models
21. Ocean and Ocean Craft
22. New Wood Flooring
23. Space Wrench
24. Mohole and Sealab II
25. Introduction to Man and the Group
26. Microwave Energy
27. Population and Education
28. Science and Technology
29. Satellite "Score"
30. Careers
31. President Johnson's Statement

Figure 6. Physical layout of the mock-up validation.



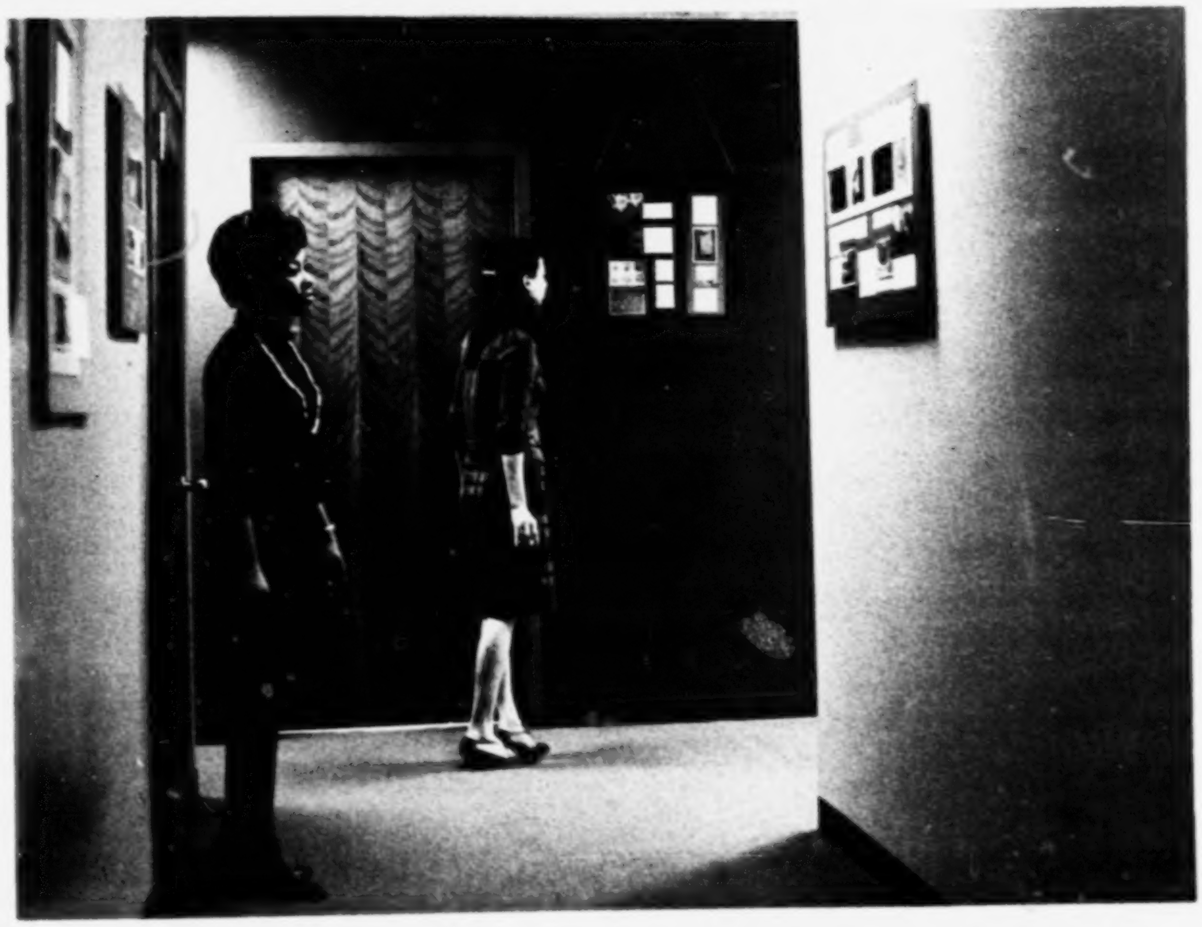


Figure 7. Photographs of study subjects viewing mock-up introductory and subarea panels.

seven basic questionnaires. All study subjects returned to the conference room for testing. This basic procedure was followed for all three high schools tested in the mock-up validation phase of this study.

At a later date, the three high schools sent IQ scores for all of the students who had participated in the study.

Phase 2. Mock-up variation testing. Data analysis of the mock-up validation phase indicated that the mock-up was a valid replication of the actual exhibit. Therefore, it was feasible to investigate the significance of several design variables by manipulating them within the mock-up structure.

The exhibit topics used in this mock-up variation study were limited to four because of time restrictions on testing and subsequent data analysis. The four general topics were: Amplification, Substitution and Extension; Fish Sounds; Chrysanthemums; and Cosmic Rays.

The decision regarding how many and which variables should be studies by means of mock-up variations involved weighing cost, time, and potential contribution. Based on these considerations, it was decided to investigate the following: 1) the effects of an aural presentation of the exhibit material versus a textual presentation; 2) the effects of a reduction in irrelevant and redundant text material combined with a reduction in reading difficulty; and 3) the contribution of the visual illustrations that accompanied exhibit textual content.

The design variables of sound presentation versus textual presentation were the two main breakdowns within this experimental design. The variables of content and visuals were varied within these categories. The aural presentation consisted of a monitor reading the text to the students. The textual presentation consisted of the students reading the texts themselves.

Variations within content consisted of presenting the text as it appeared in the original exhibit, called FULL CONTENT, as opposed to SKELETON CONTENT, e.g., texts which had been shortened in number of words along with a reduction in reading difficulty. Members of the project staff rewrote the original exhibit text many times before creating acceptable "skeleton" texts. An acceptable text had to incorporate all the ideas presented within the full text, be shorter in overall length, and be reduced by at least one readability grade level (as determined by the 1948 Revised Flesch Readability Formula). Table 10 shows the total number of words and Flesch grade levels for both the full text and skeleton text presentations of individual topics within each area.

TABLE 10

Summary Table of Total Number of Words and Text Grade Levels For Full and Skeleton Texts Used in the Mock-up Variation Testing Phase

	Full Text # Words	Skeleton Text # Words	Full Text Grade Level	Skeleton Text Grade Level
Amplification, Extension, and Substitution				
Introductory Paragraph		29		11.7
Amplification	151	89	11.4	9.1
Extension	115	83	8.7	7.8
Substitution	142	95	10.4	8.5
	408	296	$\bar{x}10.2$	$\bar{x}8.5$
Fish Sounds	368	179	9.5	8.6
Chrysanthemums				
Phytochrome	124	61	10.7	9.6
Mutations	563	109	10.2	9.5
Radiation - Food and Insects	277	120	9.7	8.2
	964	290	$\bar{x}10.2$	$\bar{x}9.1$
Cosmic Rays				
Cosmic Rays	148	68	8.9	7.8
Theories	61	53	13.8	12.8
Radio Telescope and Radio Astronomy	585	211	10.8	8.9
	794	332	$\bar{x}11.2$	$\bar{x}9.8$

The design variable "visuals" consisted of the pictures obtained from the exhibit which had been incorporated into the mock-up. The two variations were based on viewing pictures along with the text as opposed to not having any visuals to view at all.

Eight experimental groups, consisting of 10 subjects each, were used in this testing phase. The 80 subjects were junior and senior level students obtained from a local high school. Half of these students had strong science backgrounds, and half had a minimal background in science. The "science" students were distributed, randomly and equally, among the eight groups and, likewise, for the "nonscience" students, so that there was a total of 10 students (five "science" and five "nonscience") in each of the eight groups.

Since a list of participating students' names and respective science and nonscience classification was given to the project staff prior to testing, students were assigned to experimental groups before their arrival. Detailed plans of mock-up viewing and testing schedules were made by the project staff prior to the testing day. Simultaneous testing of 80 students assigned to eight different experimental groups tends to be a complicated process. For this testing, it was necessary to use five testing monitors, each of whom had several assignments. Prior to the actual testing, the staff monitors practiced their assignments in a "dry-run" in order to insure that the time schedules were realistic and that the actual testing would go smoothly.

The testing procedure was as follows: Upon arrival, students were given their group assignments and group number identification tags to wear. The staff supervisor introduced five testing monitors and gave a general talk about the exhibit, the purpose of the mock-up and what the eight experimental groups of students would be doing. Each group was then dealt with according to the time schedule. Refer to Appendix K for a diagram of the mock-up variation layout.

Listed below are the experimental groups and mock-up variations performed on each. Groups 1 through 4 were the nonvisual groups in this experimental design. At no time did they view the exhibit mock-up.

Group 1 - Read the full text.
Group 2 - Read the skeleton text.
Group 3 - Heard the full text.
Group 4 - Heard the skeleton text.

Groups 5 through 8 were exposed to the four panels presented in the mock-up containing the visuals.

Group 5 - Viewed the mock-up with full text and visuals.
Group 6 - Viewed the mock-up with skeleton text and visuals.
Group 7 - Viewed the mock-up with visuals and heard the skeleton text.
Group 8 - Viewed the mock-up with visuals and heard the full text.

Each group was tested immediately after exposure to either the textual content or the mock-up. All subjects were administered the complete background, interest, and attitude questionnaires. Only items relevant to the four exhibit areas covered in the mock-up phase were selected from the four knowledge measures. (Note the circled item numbers within the four knowledge measures in Appendixes G through J.) The high school also provided IQ scores for the 80 participants in this testing phase.

Summary

This completes the discussion of methodology. As the reader may have noticed, the variety of techniques used in this study precludes neat categorization. However, the essential parameters consisted of the study subjects and the casual subjects. Within the study group, time was controlled by assigning subjects to MIN, MAX, or CONTROL conditions. Also within the study group, systematic variations were introduced relating to education, age, sex, socioeconomic level and science, nonscience background.

Both the study and casual groups were administered comprehensive tests covering all aspects of the exhibit, i.e., knowledge, interest, and attitude. Study subject prior knowledge was measured by CONTROL group testing; casual viewer subject prior knowledge was measured by administering pretests.

The exhibit was replicated in a simple, two-dimensional mock-up; and the study group design was repeated but with high school students only. A portion of the mock-up was then used to study exhibit variations, i.e., sound versus reading, simplified text versus original and visuals versus no visuals.

Other kinds of data were collected to supplement those obtained by the paper-and-pencil tests. Included were video tape observation of casual viewer behavior at selected displays, and time data (i.e., number stopping and length of stop) on casual viewers at each display area in the exhibit. A readability analysis of the exhibit text materials, an analysis of typesize, and an analysis of the number of dynamic and static displays were also carried out.

These are the essential elements in the study. In the next section, the results of their manipulation and analysis will be presented.

RESULTS

Biographical Data

Background information was collected for all subjects at both exhibit sites, and included data on age, sex, student status, number of science courses, etc. (The questionnaire itself is shown in Appendix D.)

Biographical data of this sort is helpful in a general way only to make gross comparisons. To be useful in terms of analysis of exhibit effectiveness, it must be taken into account systematically in the design of the study. In the present instance, the factors considered most important were included as part of the audience variables, i.e., science-nonscience, male-female, socioeconomic level, educational level, etc. The effects of the many remaining variables that might have been measured, but were not, are considered as sources of random variance and are assumed to distribute themselves randomly across all conditions. It is obvious, however, that variables not considered in the study may, in fact, be more related to measures of effectiveness than the ones selected. This is only to say the obvious -- namely, that the composition of an exhibit audience is a tremendously complex element to consider in any exhibit study, and furthermore, that no claim to knowledge concerning exhibit design and effectiveness variables will ever by any better than our understanding of the audience variables.

The fact that we know relatively little about our audience in terms of how they learn, form attitudes, change interest, or become motivated are weaknesses that plague all of our educational media houses. However, this does not provide a license to ignore the audience or to treat it simplistically. While the present study cannot claim to have more than scratched the surface, it did attempt to deal with the audience in a systematic way. Exhibit studies must continue to do this if there is any hope at all of being able to design exhibits for an audience rather than at an audience.

For those readers interested in a summary of background characteristics and biographical comparisons between the casual and study groups for Los Angeles and Chicago, a table of data and a discussion of these data are contained in Appendix L. The individual items of biographical data relevant to a particular analysis will be discussed in the context of that analysis.

Interest Results

The posttest interest index consisted of two parts. Viewers were first asked if this exhibit stimulated their interest in a particular topic or area. If they answered "yes," they were asked to write the topic or topics which they found most interesting. If they answered "no," they were asked to go on to the second part of the questionnaire.

It is generally acknowledged that people tend to give "acceptable" answers to questions when surveyed, a phenomenon sometimes referred to as the courtesy bias. An effort was made to circumvent this bias in the interest index by asking "yes" respondents to recall and write the exhibit topic they found particularly interesting. It was assumed that people who could recall a topic were more than likely interested in it. Those who couldn't were perhaps just making an effort to be pleasant.

Table 11 shows the answers for all Chicago posttest groups to this first question. Note that the percentages in the category -- "Yes" Group Who Named Topic -- are based on the number of total yes answers only.

TABLE 11

Answers of Chicago Posttest Groups to Interest Question: "Did the Exhibit Stimulate your Interest in a Particular Topic?"

	No	Yes	% of "Yes" Group Who Named Topic
Posttest Casual Viewer N=303	56%	44%	93%
High School Min Group N=58	31%	69%	100%
High School Max Group N=55	22%	78%	100%
College Min Group N=19	42%	58%	91%
College Max Group N=19	32%	68%	100%
Adult Min Group N=13	31%	69%	100%
Adult Max Group N=13	54%	46%	100%

Only 44 percent of the posttest casual viewer group reported that the exhibit was stimulating. However, 93 percent of these people were able to name a topic. Fifty-six percent of this group were willing to state that the exhibit did not interest them. This could be considered a conservative figure in view of the courtesy bias. Moreover, the total number of people in the casual viewer group should be 320 rather than 303. The missing 17 people would not answer the first question, which might be interpreted as an unwillingness to commit themselves in a negative direction.

With the exception of the college MIN and adult MAX groups, the paid study group data in Table 11 show a generally positive response which is supported by the data on naming the topics. More college MIN people answered "yes" than "no" to the first question although the percentages are fairly close. Since the total number of people in this group is small, the 91 percent naming a topic means that only one person out of 11 "yeses" was _not_ able to recall a topic. In the adult MAX group, seven people state that the exhibit did not stimulate their interest, while six people answered yes and could recall a topic. Because the total number of people in this group is small, no general conclusions should be drawn from these percentages.

The data from the first part of the posttest interest index indicate that a large proportion of the high school and college study subjects reported the exhibit to be "stimulating," while more than 50 percent of the casual viewers did not or were unwilling to commit themselves. The amount of exhibit viewing time is probably a factor here. Most casual viewers tend to spend relatively short periods of time in any one display area unless the topic appears interesting at "first glance." They often move quickly throughout an entire exhibit without stopping at any one place for more than a few seconds. When the minimum and maximum exposure groups for high school and college subjects are examined separately, the MAX groups found the exhibit more stimulating than the MIN groups. Although, this is not a statistically significant difference, it does suggest that the total amount of viewing time may be positively correlated with an individual's postviewing interest level. The fact that the study groups were especially selected and paid for their participation may also have influenced this group toward a more positive level of interest.

The second part of the interest index was administered to both pretest and posttest subjects. This consisted of five short descriptions of the exhibit areas. (See Appendix E.) Subjects were asked to rank these areas from one to five according to how interesting they sounded. A rank of one equaled "most interesting"; a rank of five, "least interesting." Since this ranking system was not used during the Los Angeles tryouts, the results from the two testing sites are not directly comparable. Only Chicago results will be presented and discussed here.

A mean rank for each exhibit area was computed within each experimental group. For example, the rankings for the high school MAX group were listed for the area, Man and His Earth, summed and divided by the total number of subjects.

Figure 8 shows the pretest and posttest casual viewer mean interest ranks for the five areas. Note that the scale has been drawn so that the lowest ranking of 1.0 is at the top of the figure to indicate highest interest.

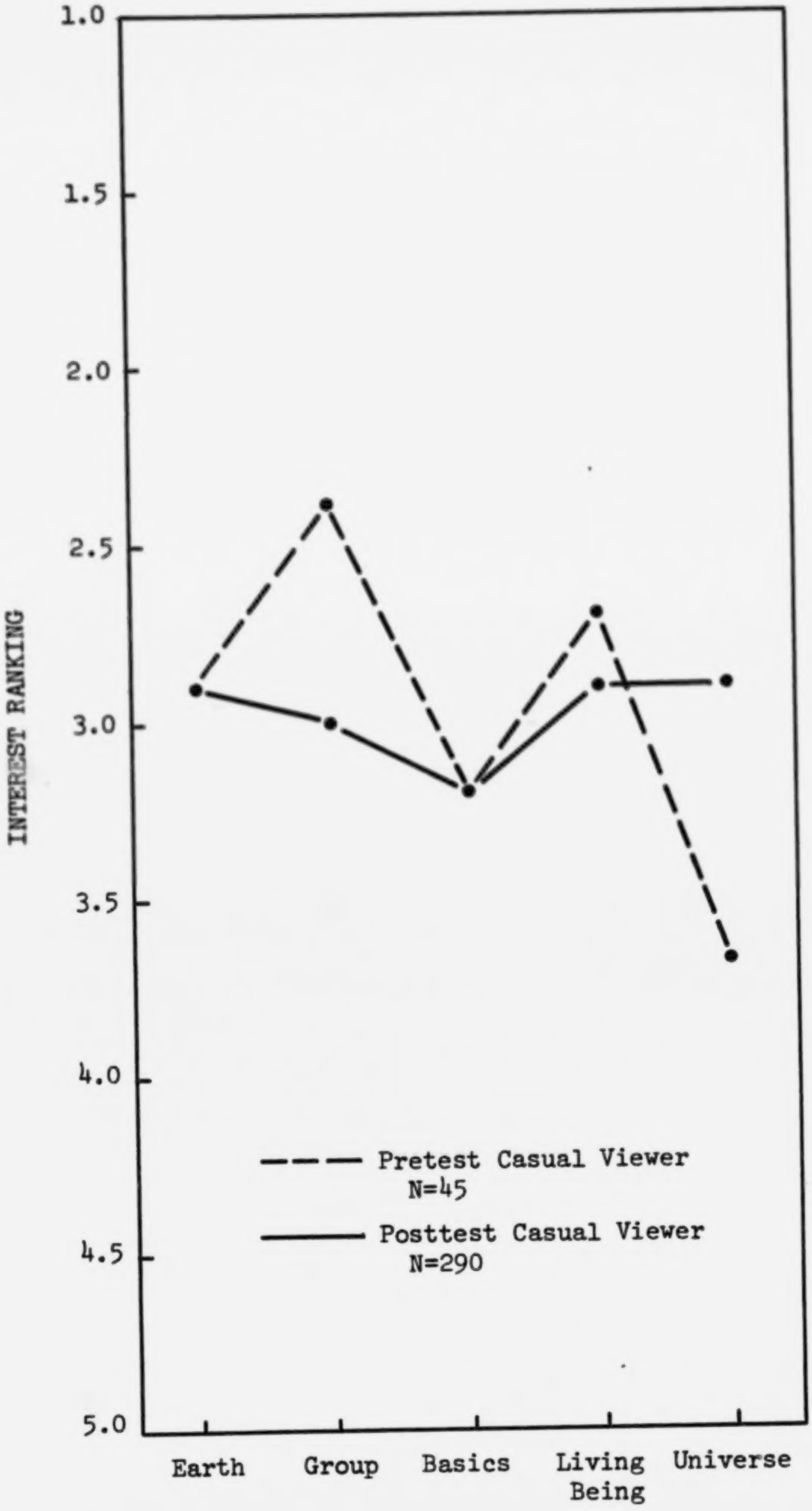


Figure 8. Chicago casual viewers' mean interest ranks of exhibit areas.

The pretest casual viewer group had not seen the exhibit when they filled out the questionnaire. It appears that this group had definite opinions about what would be interesting and high expectations for two of the areas, Man and the Group and Man, the Living Being. This group had very low expectations for Man and the Universe (3.7). The posttest casual viewer group shows a leveling of interest across the five areas with a difference of only .4 mean points between the highest and lowest rated items. The similarity of the pretest and posttest mean ranks for Man and His Earth and Man and the Basics indicates that these areas when viewed met the expectations of the pretest group. Man and the Group does not appear as interesting to the posttest group as was expected on the basis of the results of the pretest group. The posttest casual viewers ranked Man, the Living Being slightly lower than the pretest group, but this difference is not significant. The disparity between the mean ranks for Man and the Universe indicates a noticeable improvement in the viewer group when compared with the low expectations indicated by the nonviewing group.

Figure 9 shows the mean interest rankings for the MIN, MAX, and CONTROL groups within each age grouping. Although the results presented are not clearcut and can be interpreted only in general terms, there are definitely similar patterns for the MIN and MAX groups within the high school and college age groups. Not only are the MIN and MAX patterns similar in each age group, but the overall patterns are the same across the two age groups. There are some differences within each experimental group, but the overall patterns are the same. The adult interest rankings do not follow any one pattern. Neither of the adult MIN nor MAX groups look like each other or the high school and college groups. When one considered the heterogeneity in the makeup of the adult group in comparison to the relative homogeneity of the school age people, this is not surprising. High school and college students have many things in common, whereas the adult groups varied widely in age, education, socioeconomic level, etc.

The three CONTROL group patterns are quite interesting. The college and adult groups look similar here with the same expectation patterns. Note that these patterns are replicated in the pretest casual viewer group. (See Figure 8.) All three groups expect the areas Man and the Group and Man, the Living Being to be most interesting. Notice the disparity between the expectation level for Man and the Group and the obtained interest level of both college and adult MAX groups. Both CONTROL groups expected Man and the Group to be the most interesting while the two MAX groups rated it as least interesting after having viewed the exhibit.

These results on interest levels reveal that interest can be measured and related to audience variables and viewing conditions. Such data can provide valuable information for the exhibit designer. Low interest areas would need to be given special attention in terms of lighting, models, sound, etc. Less attention could be given to high interest areas. The designer can capitalize on the existing

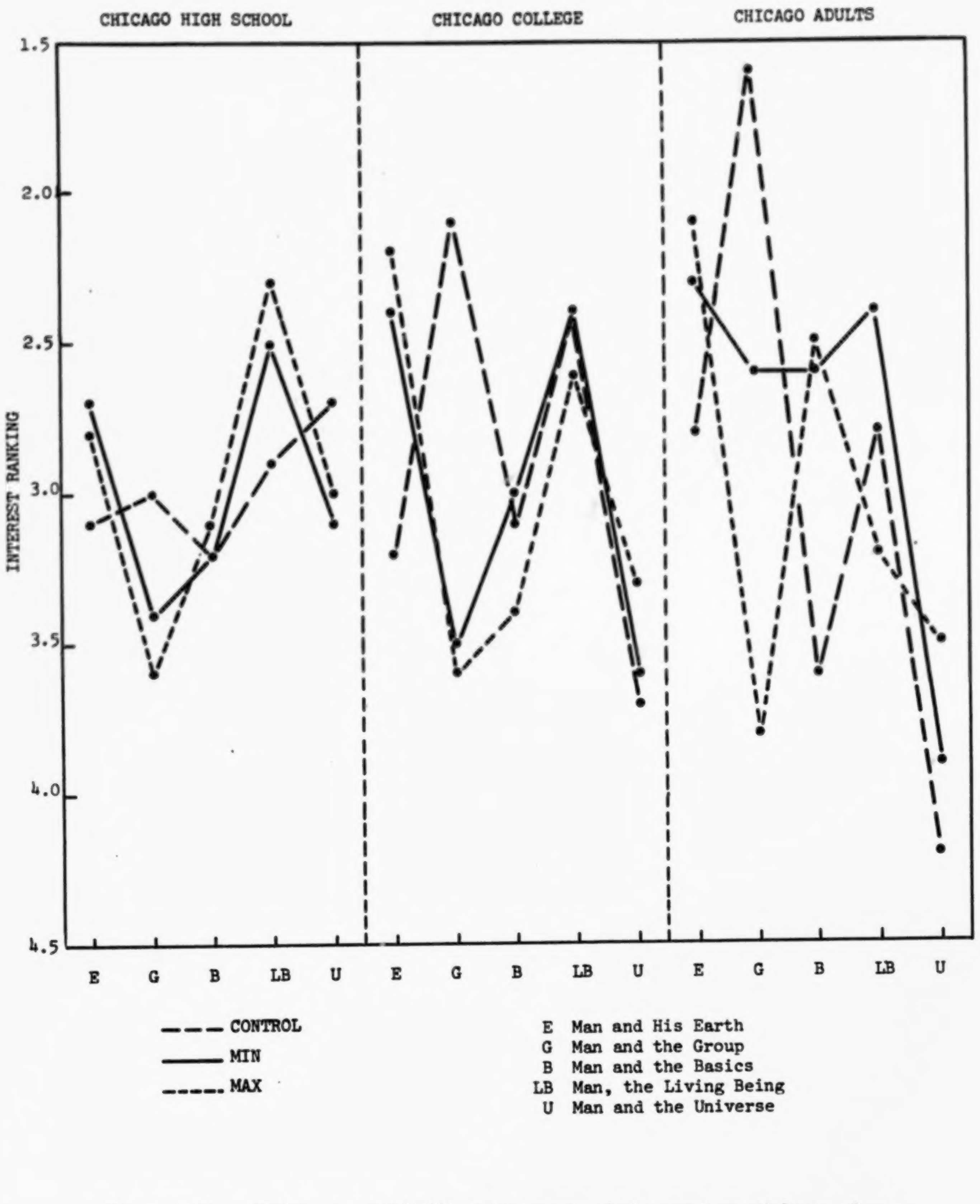


Figure 9. Chicago exhibit area mean interest rankings for MIN, MAX, and CONTROL within each age group.

interest, something that was not done in the Vision of Man Exhibit, particularly for the area Man and the Group. Thus, exhibit costs could be distributed more rationally, with the result that funds would not be invested in getting viewers "interested" in topics in which they are already interested.

Interest Conclusions and Discussions

The interest measures gave rather gross results about the level or amount of interest created by the exhibit and its subareas. Each item will be examined and difficulties noted.

A. Question: Did exhibit stimulate your interest in any particular topic?
   1. Subject to courtesy bias.
   2. If subject answered "yes," worth only one point.
   3. If subject answered "no," worth zero and there was no way to recoup; that is, no effort was made to find out why it didn't stimulate viewer's interest.

B. Name the topic or topics that interested you if you said "yes." These answers, which often mentioned a specific subarea, were scored only for the large area to which they belonged.
   1. Five areas -- therefore only five points possible.
   2. A person could name five topics within an area and still get only one point.

C. Rank each area (descriptions given) from one to five.
   1. The ranking technique is of merit for evaluating one exhibit because it does show the relative popularity of each exhibit area. But it was not possible to get an overall interest score with this ranking procedure [i.e., anyone who knew how to rank (and many didn't) automatically scored 15 points].
   2. The only useful data from ranking was the different patterns of interest for CONTROL versus MIN and MAX and CASUAL.

That several of the interest scores showed consistent patterns strongly suggest, even with the problems noted, reliable indications of interest levels were being obtained.

Attitude Results

The attitude questionnaire was designed to measure both overall attitude change as a result of viewing the exhibit and change related to specific exhibit objectives.

Figure 10 presents the mean attitude scores for the Los Angeles and Chicago subjects. These scores were calculated by summing the total attitude scores for each group and dividing the sum by the total points possible. For example, the Chicago MIN group consisted of 90 people whose individual attitude scores summed to 2218 points. Since the highest possible score on the questionnare was 38 points, the total possible points for the group equaled 3420 (38 points x 90 people). The results were that the MIN group obtained an overall attitude score of 65 percent [100 (2218 divided by 3420)].

Notice the general flatness of both the Los Angeles and Chicago attitude patterns. It appears from these results that the amount of viewing time does not change basic attitudes. The attitudes of the pretest and posttest casual viewers from both exhibit sites are not significantly different from the study groups. Also there is no significant difference between the control groups and the MIN and MAX groups.

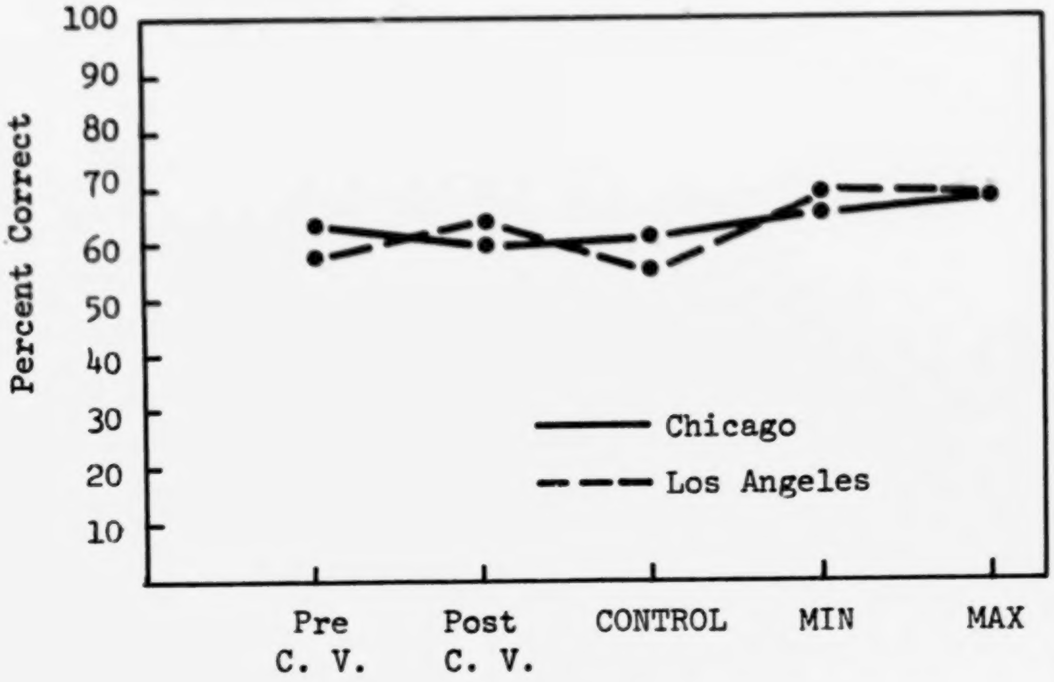


Figure 10. Overall attitude percent correct for Chicago and Los Angeles subjects.

The reader may find these results surprising. However, Figure 10 substantiates previous findings re exhibit attitude change. One effort to measure the impact of, and reactions to, an official United States exhibit at an international trade fair in Tokyo in 1957, produced disappointing results. To quote from the report:

> *We do not believe that any measurable change of an individual's attitude or opinion toward any basic question or concept ordinarily results from a single brief exposure to an exhibit, display, "message," advertisement, demonstration, theme or slogan. Long experience in attempting to measure the impact and results of advertising publicity and public relations, even after some months of continuous or repeated exposure, certainly confirms this belief.*
>
> *Even assuming that small changes of attitude or opinion toward basic questions could be consistently induced by a single exposure to an exhibit, the margin of error in the methods of measurement available to us is greater than most of the changes we attempt to measure under the prevailing conditions.* (17)

Even though the MAX groups had more than a brief exposure to the exhibit, they did not show a subsequently greater increase in overall attitude. Either a change in attitude did not occur or the measuring instrument used was not sensitive enough to record any small changes that did occur.

Another large-scale study of interest was conducted by the Institute for Sociological Research of the University of Washington on the United States Science Exhibit at the 1962 Seattle World's Fair (23). In summary of the elaborate methods employed to measure attitude change, the report states:

> *Patterns of attitude change were analyzed separately for each Pavilion hall. Results were as follows:*
>
> a. *Portions of the Pavilion produced significant changes in attitude, but the changes were of slight magnitude.*
> b. *The majority of significant attitude changes occurred in response to the "House of Science" film in Hall I. The film's general effect was to make scientists seem more academic and more eccentric. Science came to be seen as warmer and more feminine, but the public's conception of science became more vague.*

c. *Although some attitude change took place after exposure to Hall II (Development of Science), it is likely that the change came not as a reaction to the exhibits there, but instead represented recovery from changes induced by the "House of Science" film.*

d. *Attitude shifts appeared after exposure to the "Spacearium" Show in Hall III, but no obvious patterns of change were readily perceivable.*

e. *The displays in Hall IV on "Methods of Science" produced minimal attitude change. There is no evidence that they changed or clarified the public's overall understanding of the scientific method.*

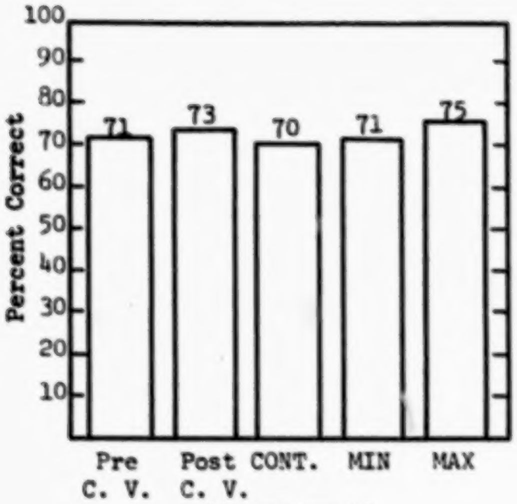
In short, an extremely elaborate and expensive exhibit had very little measurable effect on attitudes as determined immediately after exposure, and in some cases, the change in attitude that did occur was not in a desirable direction, (e.g., scientists were seen as "more academic" and "eccentric").

The results from the measurement of exhibit objectives in the Chicago testing are similar to those reported for the Seattle study. Figure 11 presents the scores for each exhibit objective. Four of the objectives were measured with items from the attitude questionnaire. Objective #5 consists of two attitude items plus items #12, 13, and 14 from the background questionnaire. The background items deal directly with interest in science, studying science, and a scientific career.

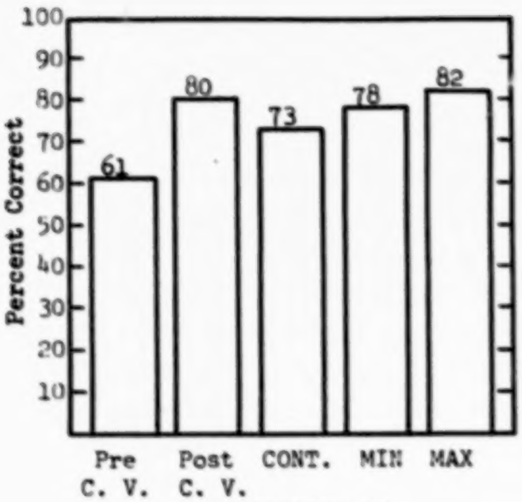
The score distribution for Objective #1, "Awareness of the Basic Achievements and Impact (technological applications) of Science," shows a small increase in positive attitudes as viewing time is increased. However, increase of this magnitude could occur by chance alone. Notice also that the pretest casual viewer and MIN experimental group obtained identical scores.

The pattern in Objective #2 shows that exposure to the exhibit did improve understanding with respect to the interaction between different sciences and between science and technology. Although these increases were not statistically significant, the pattern is in a positive direction.

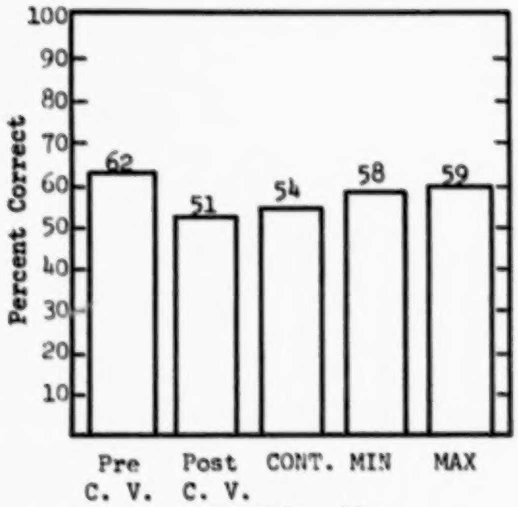
Objective #3, "Awareness of (and regard for) Relation between the Federal Government and Scientific Projects" shows the poorest percentage pattern with regard to any of the objectives. The increases among the study groups are small and the pretest casual viewer group actually scored higher than either the posttest casual viewer group or any of the study groups.



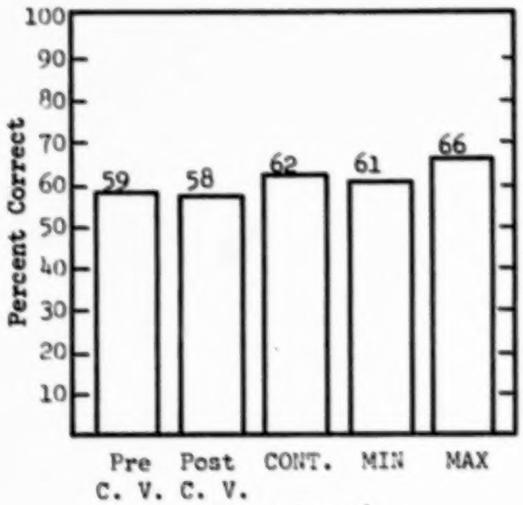
Objective #1
Basic achievements and
impact of science.



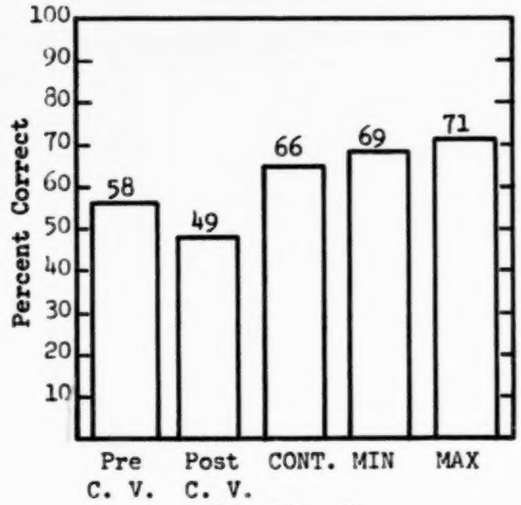
Objective #2
Awareness of the interaction
between different sciences
and between science and technology.



Objective #3
Awareness of (and regard for)
relation between Federal
Government and scientific projects.



Objective #4
Regard for scientists and
scientific projects.



Objective #5
Interest in studying science
and in a scientific career.

Figure 11. Summary of attitude percent correct for exhibit objectives.

The fourth objective, "Regard for Scientists and Scientific Projects" illustrates the general leveling pattern that occurs in many attitude studies. Although the MAX group shows the highest scores, the CONTROL group is one percentage point higher than the MIN group.

The low scores obtained by the posttest casual viewer group on Objective #5, "Interest in Studying Science and in a Scientific Career," indicates that this objective was not met with many of those in this group. The pretest casual viewer score is considerably higher but not as high as the CONTROL group. The increases from CONTROL to MIN to MAX are not significant.

A third way of examining the attitude data is in light of the science/nonscience dichotomy. Since the Vision of Man was a scientific and technical exhibit explaining basic and applied research projects, one might hypothesize that subjects who had scientific backgrounds would attain a more favorable attitude score than subjects who had nonscience backgrounds. Figure 12 shows the attitude scores for the total science and nonscience groups within the Chicago experimental viewing conditions of MIN, MAX, and CONTROL.

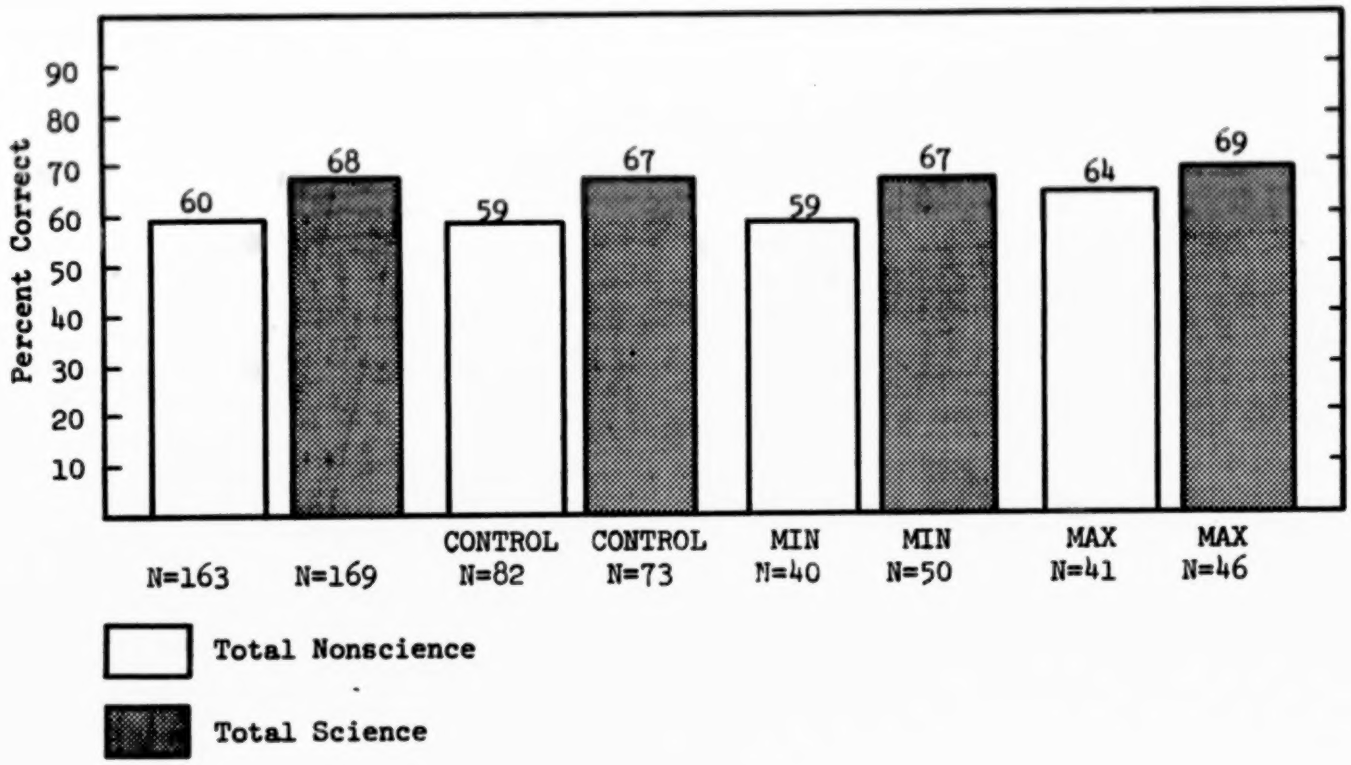


Figure 12. Percent correct, attitude scores, for Chicago science and nonscience study groups.

In every comparison, the science group achieved higher attitude scores. However, the observed differences between science and nonscience subjects are not statistically significant. Both science and nonscience scores are the same for the CONTROL and MIN groups, indicating that viewing the exhibit for a half hour did not change basic attitudes in either science or nonscience people. The MAX science and nonscience groups each show a slight elevation over MIN and CONTROL, but these differences are not significant. This suggests that unlimited viewing time may produce more favorable attitudes. This increase could, however, be due to chance factors.

The three breakdowns of attitude data have consistently indicated that the basic attitudes held by subjects are not significantly changed by viewing an exhibit as measured by the tests used. All groups tended to have similar attitude scores regardless of viewing the exhibit.

The attitude data collected in this study supports the view that basic attitudes are not easily measured and are very likely not easily changed by viewing an exhibit (20,21). One must still conclude that attitude is an extremely difficult and fragile entity to measure. Measuring instruments are usually not accurate enough to pinpoint subtle and minute changes in attitude. [A recent book, "Unobtrusive Measures" (24) is recommended as a source of interesting and creative methods for obtaining objective "belief" data.]

Who can say that no one was deeply affected by Vision of Man; what value can be placed on the influence Vision of Man may have had on one child, who later went on to become an important research scientist making important discoveries? But such speculation is a source of small consolation to those who demand more substantive evidence that exhibits like Vision of Man are worth the time and effort required. As has been noted, the purpose of this study is not to evaluate Vision of Man per se. Results such as those presented here cast serious doubts on the entire field of exhibit design and evaluation insofar as attitudes, beliefs, and opinions are concerned. Better statements of objectives, better knowledge of the appropriate stimuli for attitude change, and better measuring devices are all required.

Open-End Concept, Open-End Knowledge, and Multiple-Choice Item Results

The rationale for the development of recall and recognition knowledge questionnaires was discussed earlier, in the Methodology Section of this report. It should be remembered that open-end concept items were developed to measure viewers' recall of scientific concepts and principles. The open-end knowledge questionnaire called for recall type answers about specific factual information. Items on the multiple-choice questionnaire supplied four alternative answers and required the viewer to recognize the correct answer.

It should be noted that while recall learning is more difficult than recognition learning for a given item of information (i.e., "When did Columbus discover America?" is more difficult than "Columbus discovered America in: 1429, 1942, 1492, 2942."), open-end items per se are not necessarily more difficult than multiple-choice items. An open-end concept item, for example, covering a general principle could be quite easy, while a multiple-choice item covering a small bit of detailed information could be very difficult. In general, for the tests prepared for this study, open-end concept items were easy, open-end knowledge items were difficult and multiple-choice items ranged between easy and difficult, and with a tendency toward the easy side.

For this analysis, the items from the three questionnaires were examined in terms of topic similarity. Most of the exhibit topics did not have comparable items from each test. A total of eight exhibit topics were measured by at least two of the instruments.

Table 12 shows the Chicago high school MIN and MAX groups percent correct for the comparable items on each topic.

TABLE 12

Chicago High School MIN and MAX Groups Percent Correct
On Comparable Concept, Knowledge, and Multiple-Choice Items

Topic	Open End Concept	Open End Knowledge	Multiple Choice
Project Mohole	75%	34%	79%
New Wood	69%	47%	61%
Fish Sounds	94%	76%	
Desalination	82%	45%	
DNA-Genetic Code	85%		89%
DNA-Earliest Form		21%	52%
Polymerization		15%	50%
Laser Beam		66%	87%
Technology		59%	68%

These results generally support the rationale behind the effort to tap different levels of learning. Most students were able to recall general scientific principles but were not as adept at recalling the specific facts supporting the principle. However, they could recognize these specific facts when four choices were given in a multiple choice format. The open-end concept and multiple-choice percentages are considerably higher than the open-end knowledge percentages on all topics. These results are supported by similar data from the Los Angeles testing.

There were not enough comparable items for each topic to determine whether the differences, due to question format, were statistically significant. The results strongly indicate that exhibit learning may be attained at several levels. It is suggested that further studies into exhibit effectiveness should more thoroughly investigate differences in the levels of learning by designing several comparable test items within the framework of recall concept, recall knowledge and recognition. If the preliminary results presented above are validated by other studies, they may well have an effect on future exhibit designs. Perhaps before an exhibit is built, designers will decide exactly what type of information they want a viewer to learn from an exhibit. Once this decision is made, subsequent decisions regarding information emphasis, presentation technique, and sequencing will be based on a more rational criterion. It has also been suggested that exhibits aiming at more than one audience level should deliberately stratify the content so that children, or those only mildly interested, or persons of low IQ, reading ability, etc., can get the main idea, while those who want more detail can find that too, but in a different format and location. The results given here would even suggest a triple segmentation, with large, easily read text for the "open-end concept" level, smaller, more complex text for the "multiple-choice" level, and smaller text yet for the "open-end knowledge" level. If such an approach were to be used consistently, viewers would learn the "code," and be less likely to engage in random scanning behavior, as they do now.

Exhibit-Only Results

Twenty multiple-choice items pertaining specifically to the exhibit were administered to the MIN, MAX, and CONTROL groups. As noted earlier, these items were designed to emphasize knowledge that was exhibit-specific, i.e., could only be answered by those viewing the exhibit. As shown by the results in Table 13, this aim was not fully realized. Both high school and college students in the CONTROL group were definitely scoring above chance (or, 25 percent); that is, for them the exhibit-only items were tapping knowledge that had been obtained elsewhere. This is most particularly true of the college group, who answered almost half the questionnaire correctly without seeing the exhibit. Since most of the questions in the questionnaire did refer specifically to displays in the exhibit (12 of the 20 questions contain the phrase "in the exhibit" or a similar phrase), this means that high school students and college students, due to their knowledge of such things as DNA, cosmic rays, etc., can predict the displays an exhibit would use to demonstrate these topics. The adult CONTROL group, on the other hand, had a mean score of 29 percent on the questionnaire, which is very close to chance, indicating that, for them, the questionnaire was a legitimate representation of the material contained in the exhibit.

TABLE 13

Percentage Scores on Exhibit-Only Questionnaire Experimental Group and by Educational Level

	CONTROL	MIN	MAX
HIGH SCHOOL	38%	57%	64%
COLLEGE	48%	71%	78%
ADULT	29%	45%	72%

Gain scores were computed for each age group by subtracting the CONTROL group mean score from the MIN - MAX mean scores (Table 14). The gain scores for most of the groups are quite low. In the high school group and the college group, there was only a seven point difference between the MIN viewers and the MAX viewers. What they did learn, therefore, was learned for the most part during the first 30 minutes of viewing.

TABLE 14

Gain Scores over Control Groups by Experimental Group and by Educational Level, Exhibit-Only Questionnaire

	MIN	MAX
HIGH SCHOOL	19%	26%
COLLEGE	23%	30%
ADULT	16%	43%

The one exception to this was the adult MAX group. They show a 43 percent gain over their corresponding CONTROL group, as opposed to a 16 percent gain for the adult MIN group. It must be concluded that the exhibit, insofar as specific displays and design elements are concerned, had the greatest impact on adults who were encouraged to take the time to view the exhibit.

Government Items Results

The objectives of the Vision of Man Exhibit emphasize both the historical and the current participation of the Federal Government in scientific endeavors. All subarea displays in the exhibit presented information about research projects that have been supported and/or accomplished by Federal scientists. The accelerating requirements for scientists, engineers and technicians were presented in a chart showing a ten-year projection of career opportunities available with the Federal Government. At the end of the exhibit, pamphlets about Federal Government career opportunities were given to exhibit viewers. In short the exhibit was built and paid for by the Federal Government; and the exhibit objectives placed great emphasis on increasing viewer's knowledge about, and interest in, the role of the Government in scientific research.

In order to evaluate these important Government oriented objectives, several recall knowledge questions were developed within the open-end knowledge questionnaire. These items were administered to all study subjects and casual viewers, (i.e., each of the eight casual viewer test sets incorporated these items). The items and correct answers are as follows:

1. "Identify the nation's largest employer of scientists, engineers, and technicians." Correct answer: Federal Government
2. "What 'partnership' was described by this exhibit?" Correct answer: Between science and the Federal Government
3. "Who paid for this exhibit?" Correct answer: Federal Government

The scores for these three items were summed and divided by the total points possible to obtain an overall percent correct. Figure 13 shows a "Government item learning continuum" for the five experimental groups.

The pre and post casual viewer groups and the experimental CONTROL group obtained similar low scores for these Government items. The posttest casual viewer group scored one percentage point higher than the pretest group and two points higher than the CONTROL group. These differences are not significant. The increased knowledge shown by the MIN and MAX group percentage indicates that there is a relationship between viewing time and knowledge gained about the Government. However, the MIN and MAX scores are not high when considered against a 100 percent correct criterion; less than half of the MAX group was able to recall the correct answers to these questions.

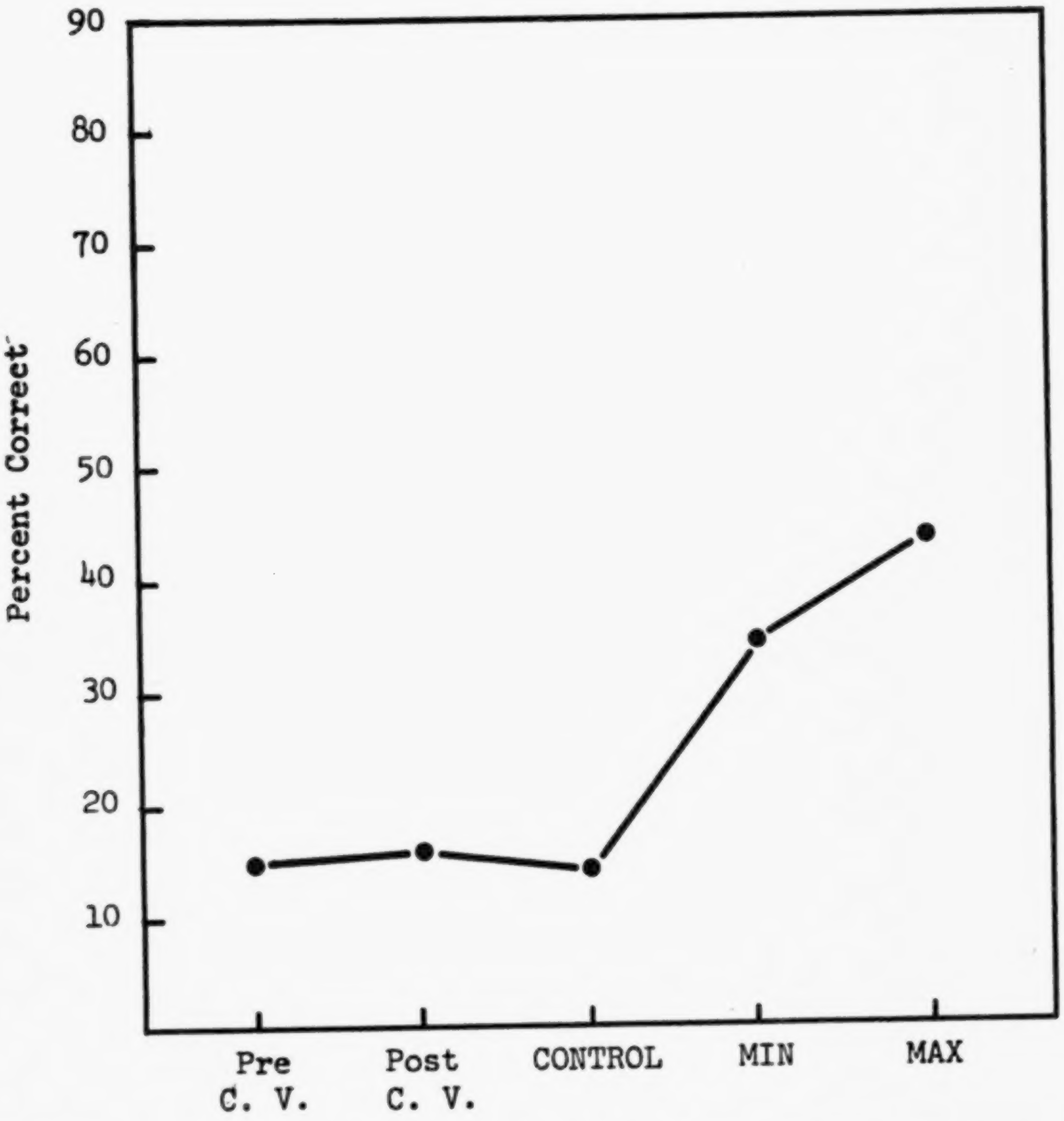
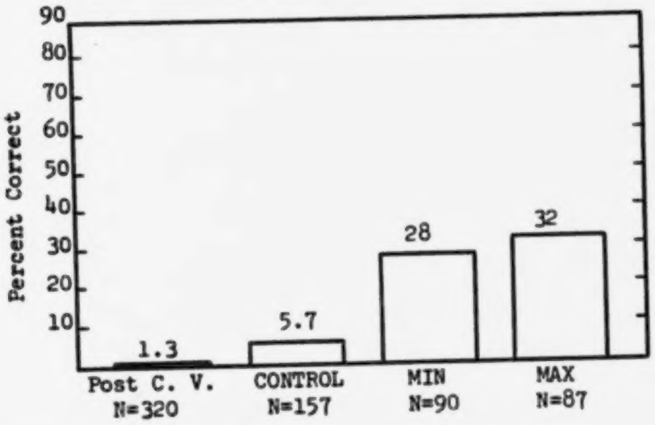


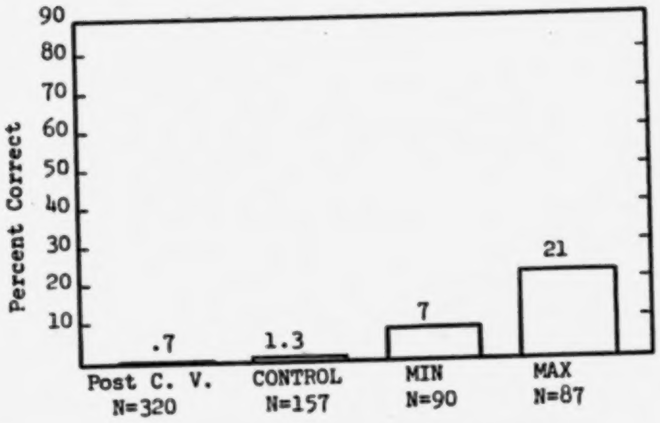
Figure 13. Percent correct for Chicago groups on three recall government items.

Since the general Government learning continuum was based on the sum of these item scores, individual item differences were obliterated. Figure 14 presents the overall percent correct for each of the three items. An additional recall question, "What is the name of this exhibit?" is shown in this figure because it indicates the degree of viewers' awareness of the exhibit as an integrated and complete display. The pretest casual viewer group is not shown in this figure.

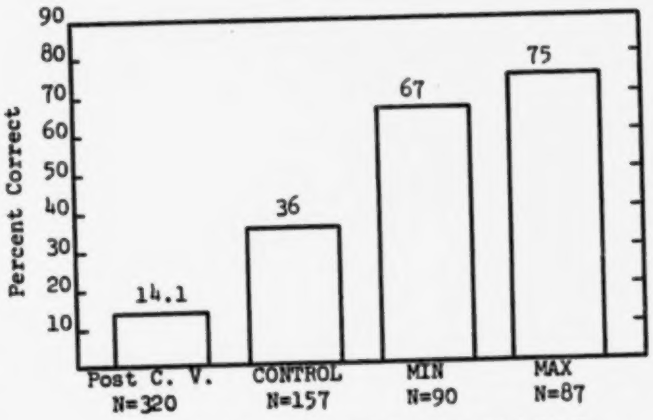
The name of the exhibit was not easily recalled by any of the groups. Only one percent of the 320 posttest casual viewers were able to state the exhibit name. Surprisingly, six percent of the CONTROL group were able to write the name. This score may be due to the publicity given to the exhibit prior to its opening in Chicago. Although the MIN and MAX groups scored higher on this item, their overall percent correct is quite low. It is apparent that exposure time did increase knowledge but not to any great degree. These data may be explained by the fact that the name of the exhibit appeared only twice within the display. A panel at the exhibit entrance stated "Vision of Man" in large lettering. The last panel, located at the exhibit exit, stated, "We hope you have



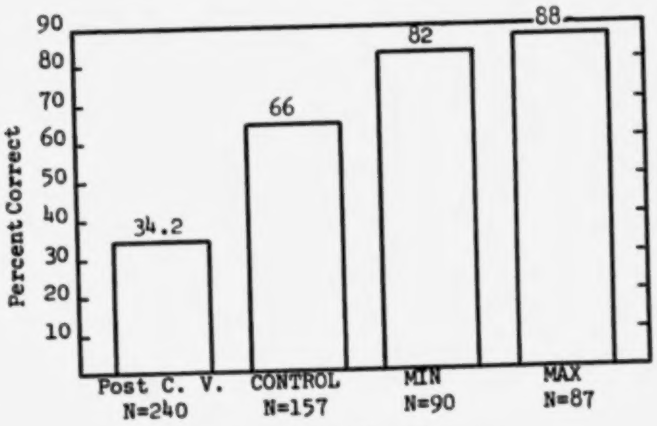
Name of Exhibit



Partnership Between Government and Science



Who Paid for This Exhibit?



Largest Employer

Figure 14. Percent correct on Government items.

enjoyed 'The Vision of Man'" in smaller lettering. The viewer thus had only two chances to learn the exhibit name. The data suggest that two chances were not enough. (When asked by an interviewer what he would tell his friends about the exhibit, one Los Angeles casual viewer stated, "I thought the DIVISION of man exhibition was a very good exhibition.")

The very low posttest casual viewer score indicates that many of these people were not aware of the Vision of Man Exhibit as a separate entity outside of all the exhibits within the Museum of Science and Industry. Two comments made by Chicago casual viewers to the above question serve to illustrate this point:

> *I'm really not sure what I looked at. I think there should be an explanation in the beginning, maybe there was and I missed it. I was walking through and really looking for another section. I was more or less lost when I stumbled into it ... needs an explanation in the beginning.*
>
> *The only thing I think needs to be improved is the starting and stopping point. I think I have come in on the middle of it or missed some of it; I'm not sure because I didn't know exactly where I was going.*

The viewer response to the name of the exhibit may be an important factor in "word-of-mouth" publicity. Very likely, Vision of Man viewers will be unable to tell their family and friends the name of the exhibit they saw.

The second distribution on Figure 14 shows viewer response to the Government item, "What 'partnership' was described by this exhibit?" The answer to this question appeared on the entrance name panel in 3/4 inch print. The statement of the partnership between science and the Government appeared only there, although all exhibit displays subtly pointed out the relationship. Less than one percent of the casual viewer group was able to answer this question. The CONTROL group percentage is similar. The MIN group did quite poorly on this item with only seven percent of the 90 people able to either recall or "discover" the correct answer. The 21 percent correct scored by the MAX group is significantly lower than any of their scores on the Government items.

This "partnership" item was developed to evaluate Exhibit Objective #3: "To increase knowledge (and regard for) the relationship between the Federal Government and scientific projects." According to the percent correct scores for this item, it must be concluded that a significant number of viewers who were tested in this study left the exhibit unaware of this relationship. This may be due to the fact that the words "Federal Government" or "Government" appeared in the entire exhibit text only 34 times out of a total of 9809 words, which is indicative of a "soft-sell" approach. Since the relationship of science and the Government was one of the five exhibit objectives, perhaps making only one direct statement about the partnership and subtly alluding to it was not enough to "sell" the idea to the audience. One Chicago casual viewer unwittingly pinpointed this problem by saying, "... when you walk in, there should be some kind of announcement or some type of eye-catching device that will enable you to know what you're getting into, because as you go further and further and you see all these different sciences, the oceanography and space, ... etc., you wonder how they do tie in." The fact that the maximum time group did so poorly on this item shows that this information was not easily learned from the exhibit even with unlimited viewing time. In view of the small amount of time each casual viewer spends in an exhibit, the fact that less than one percent of this group learned about the partnership is not surprising. However, since the data presented are based on the answers to one recall question, the results can be used only as a general indicator of the success of a soft-sell approach in this particular exhibit. Further study of the success of a hard- versus a soft-sell approach in different exhibits and with different objectives is necessary before any definite guidelines for exhibit design can be established.

The percent correct distribution on the question of "Who paid for this exhibit?" is more promising. Fourteen percent of the casual viewer group correctly answered this item. The CONTROL group did well with 36 percent. Both the MIN and MAX groups scored in the high range. The casual viewer score is low when compared with the other groups, but it is significantly higher than their responses to the first two items discussed. The response to this item indicates that even though many of the subjects were unaware of the relationship between science and the Government, they knew that the Government had paid for the exhibit. One acute Chicago casual viewer observed, "It looked like it had a pitch for the Government, but that's okay as long as they are financing it." A little old lady in Los Angeles, however, throws all of these results into question when she answered, "the Government" to the question, but then added, "Doesn't the Government end up paying for everything?" The highest percentage of wrong answers was "industry," with a number of specific companies being mentioned, usually those noted in very small labels, "Model Courtesy of ....!"

In general, many of the people tested knew that the Government was the largest employer of scientists, engineers, and technicians. Thirty-four percent of the casual viewer group were able to state the correct answer to this item. This percentage represents the highest casual viewer score on any of the Government items. However, the CONTROL group did so well on this item that one might hypothesize that a considerable proportion of the casual viewer percentage was due to previous knowledge. When the MIN and MAX percentages are compared to the CONTROL group percentage, knowledge gain percentages of 16 and 22 percent, respectively, are obtained. One can conclude that at least 22 percent of the MAX group learned that the Federal Government is the largest employer of scientists, etc., from the Careers exhibit subarea. No conclusions can be made about the percent of casual viewers who learned this information from the exhibit. It is interesting to note that out of 197 casual viewer recorded personal comments, not one person mentioned either the nation's largest employer or the career opportunities available with the Government.

MAX Results on Exhibit Areas and Subareas as Determined by the Multiple-Choice Questionnaires

As previously discussed, the total number of multiple-choice and exhibit-only items was developed from an established ratio based on the number of words contained in the textual content of each exhibit area and subarea. Therefore, the areas and subareas are fairly represented by the combined multiple-choice and exhibit-only test.

In this study, the MAX viewing group was designed to establish a "ceiling" on the level of learning that could result from viewing the exhibit. Since the Vision of Man contained many separate areas and topics within those areas, the total MAX scores on combined multiple-choice and exhibit-only items can be used as a diagnostic tool in evaluating the strong and weak subareas within each exhibit area under intensive viewing conditions. In effect such an analysis would provide data on the relative effectiveness of the various areas in the exhibit under "ideal" viewing conditions by a large and widely divergent audience, including adult, high school, and college subgroups. Weaknesses identified in such an analysis would be prime candidates for review and revision.

Figures 15A, B, and C present the Chicago MAX group percent correct for each area and inclusive subareas. Each subarea percent correct was calculated by summing MAX scores on the multiple-choice and exhibit-only items developed for the particular topic and dividing this sum by the total points possible. The figures also include the overall percent correct for each area.

The subarea percentages within Man, the Living Being illustrate the percent correct variability among the subareas. (See Figure 15A.) It is apparent that the majority of MAX viewers were able to correctly answer the item on Extension. However, the group did poorly on the items related to Amplification and Substitution, the two topics which were presented in conjunction with Extension. It may well be that this difference is due to the fact that Extension contained a dynamic model and one in which viewers could participate by pushing a button. Extension was located in the center of the display and the information was presented concisely. The models on both the Amplification and Substitution displays were static and the texts were somewhat confusing and difficult to view. The Wright Flyer percentage was based on Multiple-Choice Item #77 asking who supported early research in aerodynmaics. Only 45 percent of the MAX group selected the correct

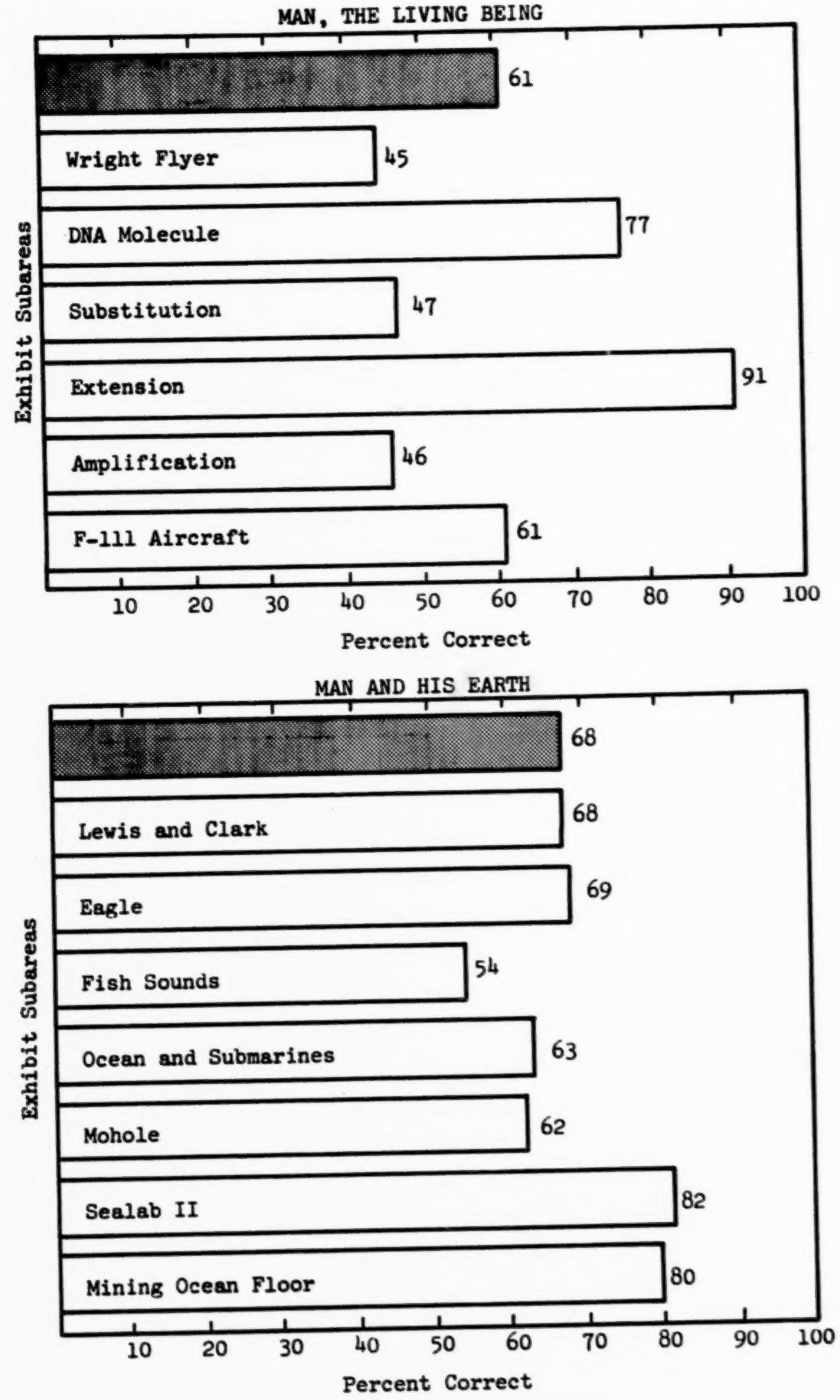


Figure 15A. Chicago total MAX group percent correct on multiple-choice items for areas and subareas.

answer, the Federal Government. This low percentage suggests that this point was not easily learned from the information contained in this display. The overall percent correct for the areas was 61 percent.

The subarea percentages for Man and His Earth show a fairly even pattern. The highest subarea percentages were within Sealab II and Mining the Ocean Floor. These subareas were presented together within the exhibit, but are separated for this analysis because they cover two different topics. The lowest subarea percentage was Fish Sounds. Out of the three items developed for Fish Sounds, multiple-choice #68 was significantly lower than the other two items and pulled down the overall percentage. Item #68 was an inherently difficult item. The overall percentage for Man and His Earth was 68 percent correct.

In Figure 15B, the subarea percentages for Man and the Basics vary from 87 percent to 49 percent correct. The items on Chrysanthemums and Mutations were easily answered by the MAX group. However, Phytochrome and Radiation of Food and Insects which were included within the Chrysanthemum display produced lower scores. The total Chrysanthemum display seemed confusing to most viewers. The topics on Phytochrome and Radiation were presented on wall panels located at the back of the overall display area. The Mutations topic covered information of the type usually found in a biology text. The topic, Stretch Cotton, was also a weak subarea according to the MAX percent correct. The overall percent correct for this area was 67 percent.

Man and the Universe appears to be the weakest area in terms of maximum learning. The overall percent correct of 52 is the lowest for any exhibit area. The two introductory boxes, Cyclotron and Goddard's Rocket, show a great disparity in scores. It is not surprising that Space Models obtained a low 51 percent correct. The models themselves are good replicas, but the signs describing each model are poorly placed and difficult to see. Often the viewer was unable to determine which sign went with which model. The MAX percentage on the movie Satellite Score was the lowest subarea percent correct within the entire exhibit. The film was presented near the exhibit exit and was easily missed by viewers. Although the film was shown continually, this display contained no distinctive or attracting qualities. Lighting also made the film difficult to see.

Man and the Group percentages are shown in Figure 15C. The MAX group obviously learned the answer to the item about Mesa Verde. Unfortunately, the Population subarea was not included in the exhibit during the Los Angeles testing, so the staff was not able to develop items for this topic. The subarea, Education, appears weak in maximum learning.

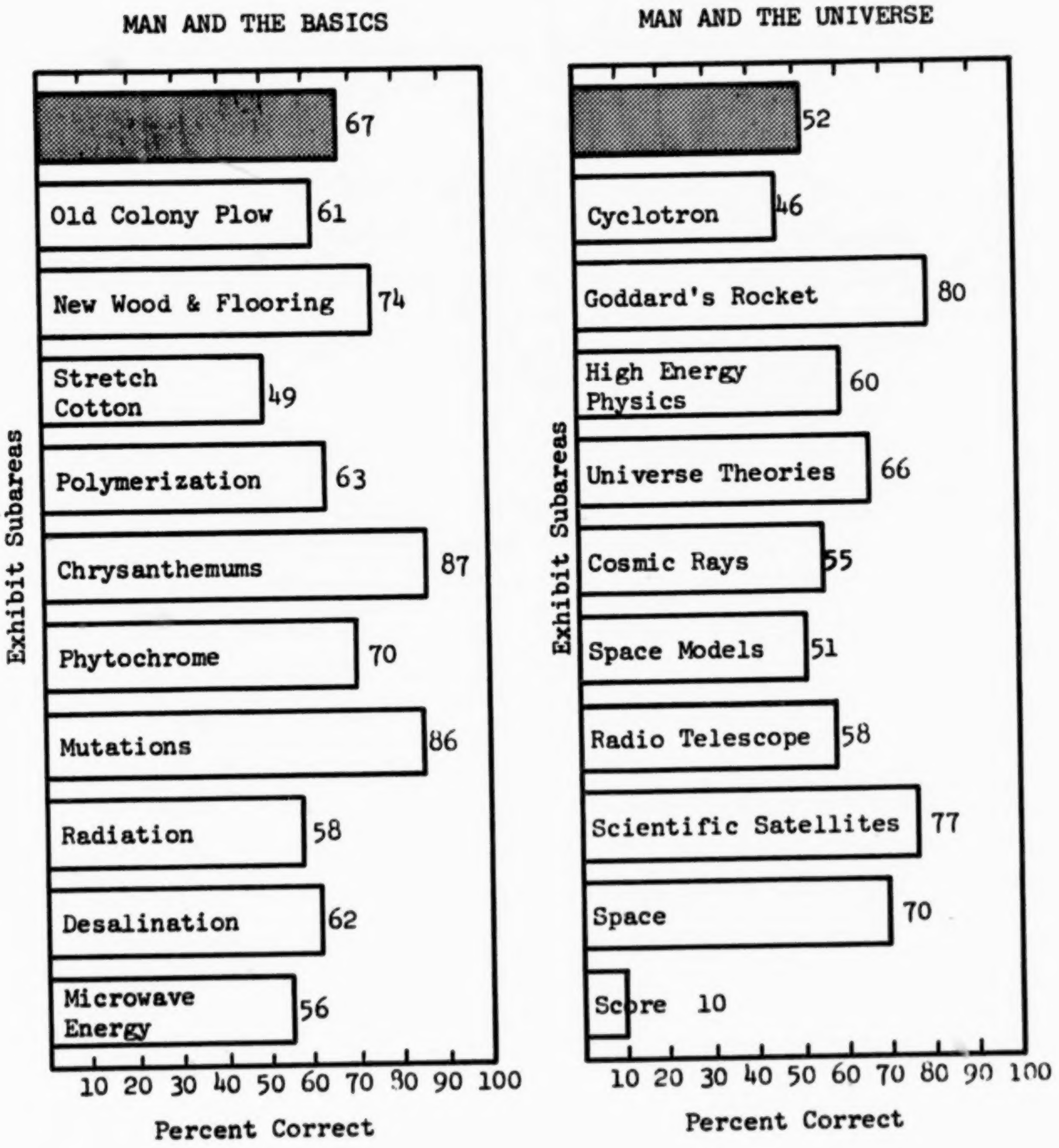


Figure 15B. Chicago total MAX group percent correct on multiple-choice items for areas and subareas.

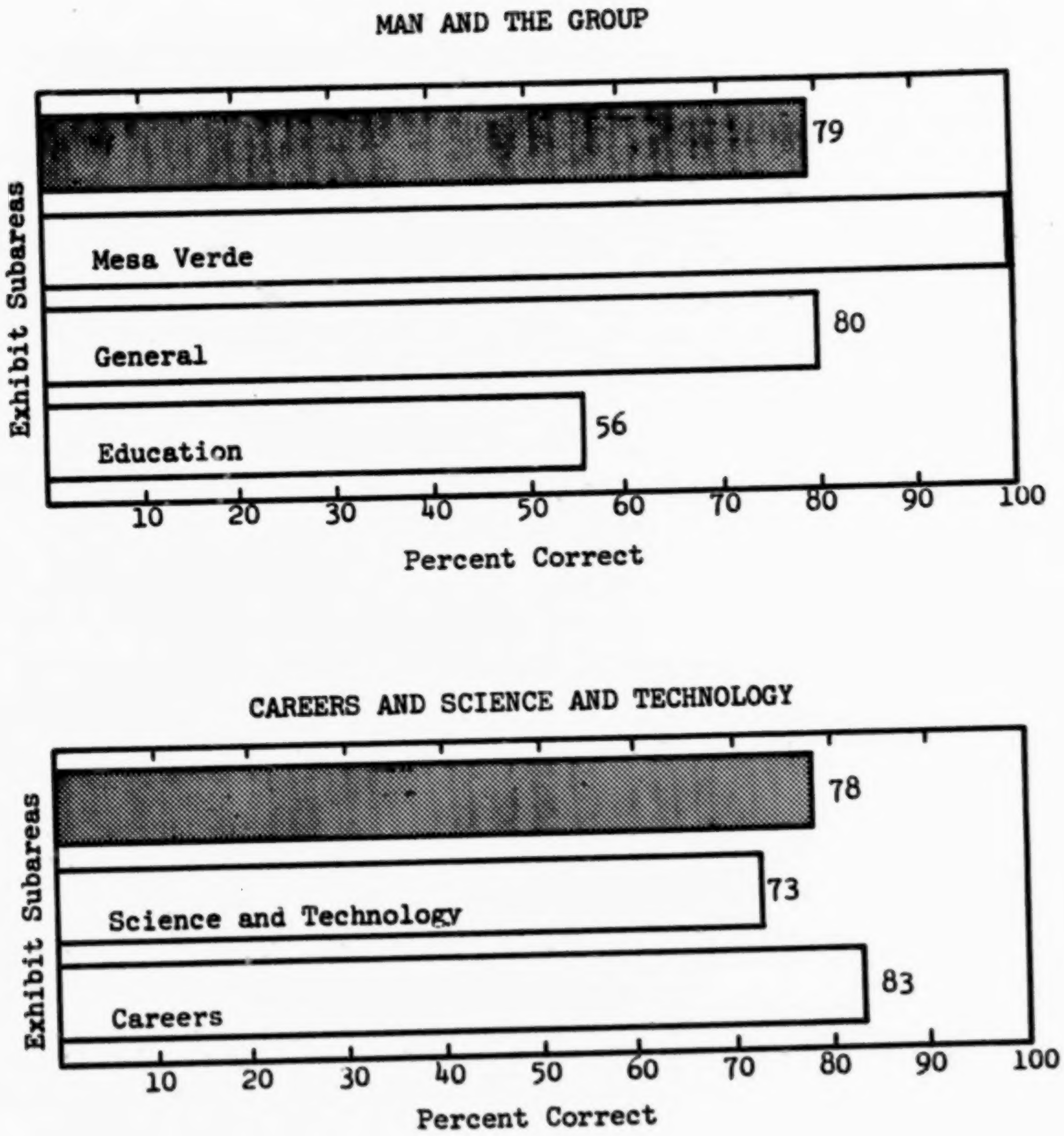


Figure 15C. Chicago total MAX group percent correct on multiple-choice items for areas and subareas.

On the average, this area obtained the highest overall percent correct. However, the nature of the content in this area necessitated the development of rather easy items in comparison with those developed for other exhibit areas. Consequently, Man and the Group is not as "strong" as it appears at first glance.

Both subareas within Career and Science and Technology obtained a high percent correct. It is apparent that the MAX group learned a great deal from this area despite the fact that the casual viewer avoided the Career subarea with alarming regularity.

These maximum learning results can be used as a diagnostic technique in determining what exhibit subareas are weak in terms of pure "teaching power." It can be assumed that if the Maximum group, who spends unlimited time in an exhibit, is unable to learn the information, the casual viewer group, who spends a small amount of time, will demonstrate very little gain in knowledge. If this technique could be used early in the process of designing and building an exhibit, methods of presentation or content could be changed before the exhibit became permanent. The exhibit mock-up validation study supports the use of such an approach. Using MAX only subjects would simplify the procedure and still provide data on the relative strengths and weaknesses of the various elements in the exhibit.

Total Knowledge Results

Each of the four knowledge questionnaires used in this study (open-end knowledge, open-end concept, multiple-choice, and exhibit-only) was developed to evaluate different types of learning that can occur as a result of viewing an exhibit. The data collected for each test suggest that there are distinct patterns in the attained level of learning for an exhibit viewer.

Previous results sections have evaluated the individual knowledge questionnaires and have pointed out these learning patterns. This section will discuss "total knowledge," which is the summation of an individual's learning across all four measures. Total knowledge percent correct is a useful graphic statistic because it illustrates the subject's gain in knowledge against the criterion of all "possible" knowledge (as measured by the questionnaires).

Results for experimental viewing conditions. Figure 16 presents the total knowledge percent correct learning continuum for all experimental groups in Chicago. The pre and post casual viewer percentages are similar, with the overall posttest group scoring only two points higher than the pretest group. The CONTROL group attained a much higher level on these combined measures. The CONTROL group, made up of those who had not seen the exhibit, scored an average of 15 percentage points higher than the post casual viewer group who had viewed the exhibit. A t test showed that this difference was significant at the <.001 probability level.

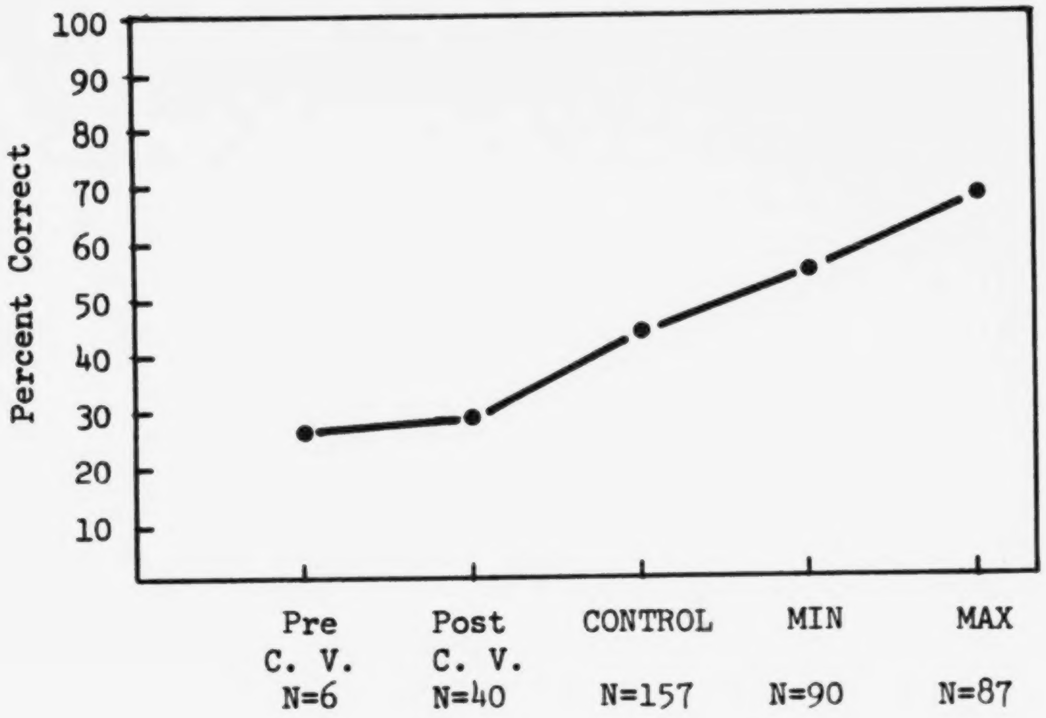


Figure 16. Total knowledge percent correct learning continuum for Chicago experimental group.

This great and surprising disparity between the two groups' total knowledge scores is difficult to explain. Perhaps with accurate subject data on IQ scores, reading level, and other standardized measures of intelligence, the reason would become clear. But it was not possible to obtain these scores for either the casual viewer group or the college and adult subjects in the CONTROL group. However, a separate analysis of relevant background items was performed for these two groups. The results indicated that the high school students in each group were similar in all respects, e.g., number of years in school, the number of science courses they had taken in school, etc. This is to be expected since the number of high school years and science courses is quite restricted. However, when the college and adult groups were singled out, several important distinctions emerged, especially in the educational area. For example, of the 159 persons in the casual viewer adult group, only 52 percent had spent any time in college, while 81 percent of the adult CONTROL group had spent time in college. Fifty-seven percent of the adult CONTROL group were college graduates as opposed to only 14 percent of the adult casual group. CONTROL group subjects had taken an average of 5.2 college science courses while the casual viewer group had taken an average of 3.7 science courses in college. In addition, the college and adult CONTROL group together had a mean of 15.4 years of schooling while the same casual viewer age groups had spent a mean of 13.9 years in school. Thirty-three percent of the total CONTROL group were currently full-time college students while 23 percent of the casual viewer group were presently in college.

These data indicate that the Chicago CONTROL subjects on the average spent more time in school and took more college science courses than casual viewer subjects. Since outside measures of intellectual competence were not available, one may only conclude that the CONTROL group seems to be better educated (as measured by years in school) than the casual group and that, at least in this respect, the casual and CONTROL groups were not drawn from the same population.

A second factor may also have influenced these scores. It can be assumed that the CONTROL group as a whole was more motivated to take the tests than the casual group. Most importantly, they were paid for their services. These people were instructed to do their best and, realizing that they were to be compared with those who actually saw the exhibit, may have taken a competitive approach to their task. In contrast, the casual group was asked to take time out of their visit to the museum to "take a test." Many of them had to be literally pushed (gently) into the testing area. Anxiety over "failing" and desire to "get out of there" are not generally felt to be good testing conditions. Guilt feelings were also in evidence, brought on by the realization that they had not really given the exhibit the serious thought that, in retrospect, it seemed to require. Perhaps the classic comment in this regard was made by the casual viewer who responded to the question, "What will you tell others about the exhibit?", with:

> *If I were going to tell my friends or family, I'd tell them to be sure and look at it carefully because you take a test after you get out.*

Other comments along these lines were:

> *To tell you the truth, I didn't pay that much attention to it. We wandered through and then when we were asked to fill out the questionnaire, we just kind of did it. We weren't really prepared for this. Could you tell me just what the screwworm fly does? It kills livestock, doesn't it?*
>
> *If I had known I was going to give an opinion or evaluation, I would have more consciously looked at the exhibit. I think I missed a lot of the ones that I was questioned on.*

In short, the CONTROL and the casual viewer group were different in at least two important respects, one having to do with their educational level and the other with their motivational and attitudinal structure re the testing situation.

A third factor that may be operating has to do not with the subjects themselves, but with the statistical manipulations required by the nature of casual viewing testing procedures. Recall that the large test designed for the experimental and CONTROL group testing was divided up into eight smaller tests for casual viewer testing. Thus, eight casual viewers equal one experimental or CONTROL subject. In this way, 320 post casual viewers became an N of 40, and 48 pre casual viewers became an N of 6. Statistically, this procedure tends to restrict the range of scores so that extremes at either end are given less weight. The influence of this factor, however, is not great, and a correction for it would probably not change the basic fact that casual viewers learned relatively little from the exhibit. The time data would serve to support this conclusion, i.e., in an average viewing time of 14 minutes, the casual viewer could not be expected to gain as much as the MIN viewers (30 minutes) or MAX viewers (unlimited, but average = 64 minutes with a range between 40 and 150 minutes).

When the experimental groups are considered alone, the total knowledge learning continuum indicates that the amount of knowledge gained increased with increased viewing time, as would be expected. However, it should be noted in Figure 16 and in the discussion to follow that the increase in knowledge is not proportionate to the increase in time. Time is at a ratio of 2 to 1 between MAX (average, 1 hour) and MIN (1/2 hour) while knowledge gain is at a ratio of 1.24 to 1. It would seem that, insofar as the Vision of Man Exhibit is concerned, and as measured by the tests prepared, knowledge gain may be reaching asymptote at approximately one hour. If one could think of it in cost effectiveness terms, the Vision of Man Exhibit is "worth" approximately one hour of time. After that, the viewer's knowledge gain is increasing at a rapidly decellerating rate. It would be most interesting to explore other exhibits to see what their "absorption" rate would be.

Several of the major audience variables will now be examined within the framework of total knowledge scores.

Results for science/nonscience categories. Figure 17 indicates that subjects with a science background (whether high school, college, or adult) scored higher in total knowledge than subjects with a nonscience background. The difference between all science groups versus the total for all nonscience groups was statistically significant at the .05 level. Note that the differences in the science/nonscience patterns across the CONTROL, MIN, and MAX groups are also in the expected order, i.e., science is always better than nonscience for each of these groups. The CONTROL science group attained the same level as the MIN nonscience group and the MAX nonscience group was only one percentage point higher than the MIN science group. It seems clear that a scientific background does influence overall learning in a scientific and technical exhibit such as the Vision of Man. Results of this sort also lend supporting evidence to the soundness of the test instruments, i.e., they seem to be sensitive to the kinds of effects that would be expected to influence them.

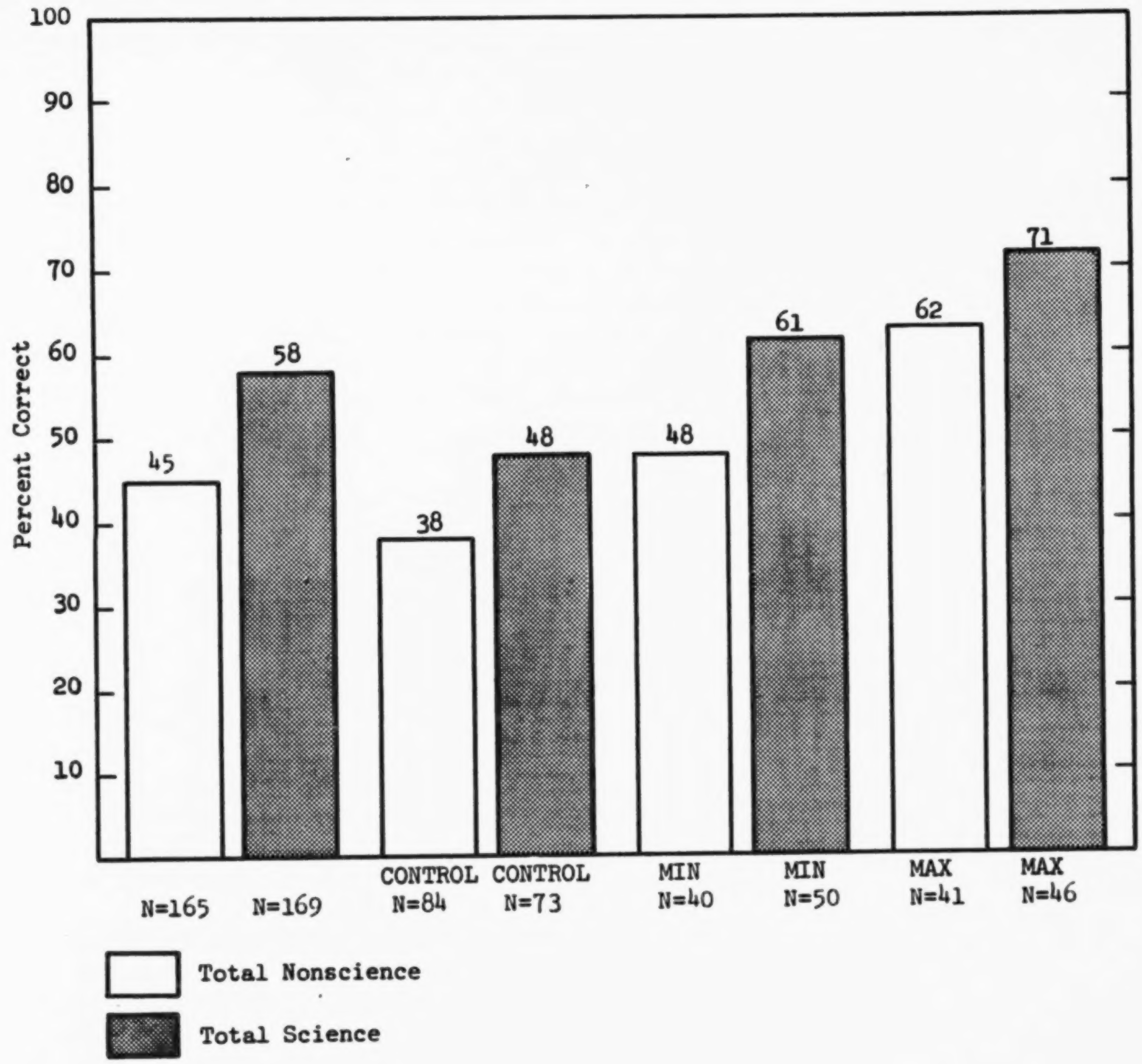


Figure 17. Summary of science/nonscience groups on total knowledge percent correct.

Results for age/education groupings. The total knowledge percent correct statistic is of use in examining the overall learning that occurs for different age/education groups. Figure 18 shows the percent correct for the total high school, college, and adult groups as well as the individual experimental conditions within each age group. Again as expected, the total college sample attained the highest overall percent correct. High school students, on the average, scored higher than adult subjects, but the adult MAX group did exceed the high school MAX group. Recall that the casual viewer posttest group averaged at 28 percent, considerably below any of the experimental age/education groups shown in Figure 18.

In each group, there is a gain in score as viewing time is increased. The adult MIN group on the whole did not learn much from the exhibit when compared to their adult CONTROL group. However, the adult MAX group shows a significant gain; the largest, in fact, of any group in this analysis.

The college MAX group score of 74 percent correct is the highest total knowledge score seen in any of the distributions. The college MIN group percentage was higher than any of the other groups, including high school and adult MAX. The college CONTROL group score was slightly higher than the high school MIN group and 12 percentage points higher than the adult MIN group.

The high school MAX group score was lower than either of the other two MAX groups. However, the high school MIN and CONTROL groups percentages were both higher than the equivalent adult groupings.

The results of this analysis of total knowledge support a basic hypothesis of this study, namely that educational level is an important audience variable, one that the exhibit designer must take into account if he is to effectively reach his audience and his objectives. The other hypothesis supported is that viewing time is correlated with learning, but that the gain in knowledge quickly drops off as time increases. While not a generalizable result, the groups used in the present study resisted breaking through the 75 percent mark in total learning. It would be interesting to see to what extent this may represent a constant function, expressed for different sized exhibits, different audiences, different objectives, etc. Perhaps such a figure could be used as a goal to "shoot" for. In programmed instruction, the specifications for the developer often read 90/90, i.e., 90 percent of the students must achieve a 90 percent score on the criterion test. What would be an appropriate figure for exhibits?

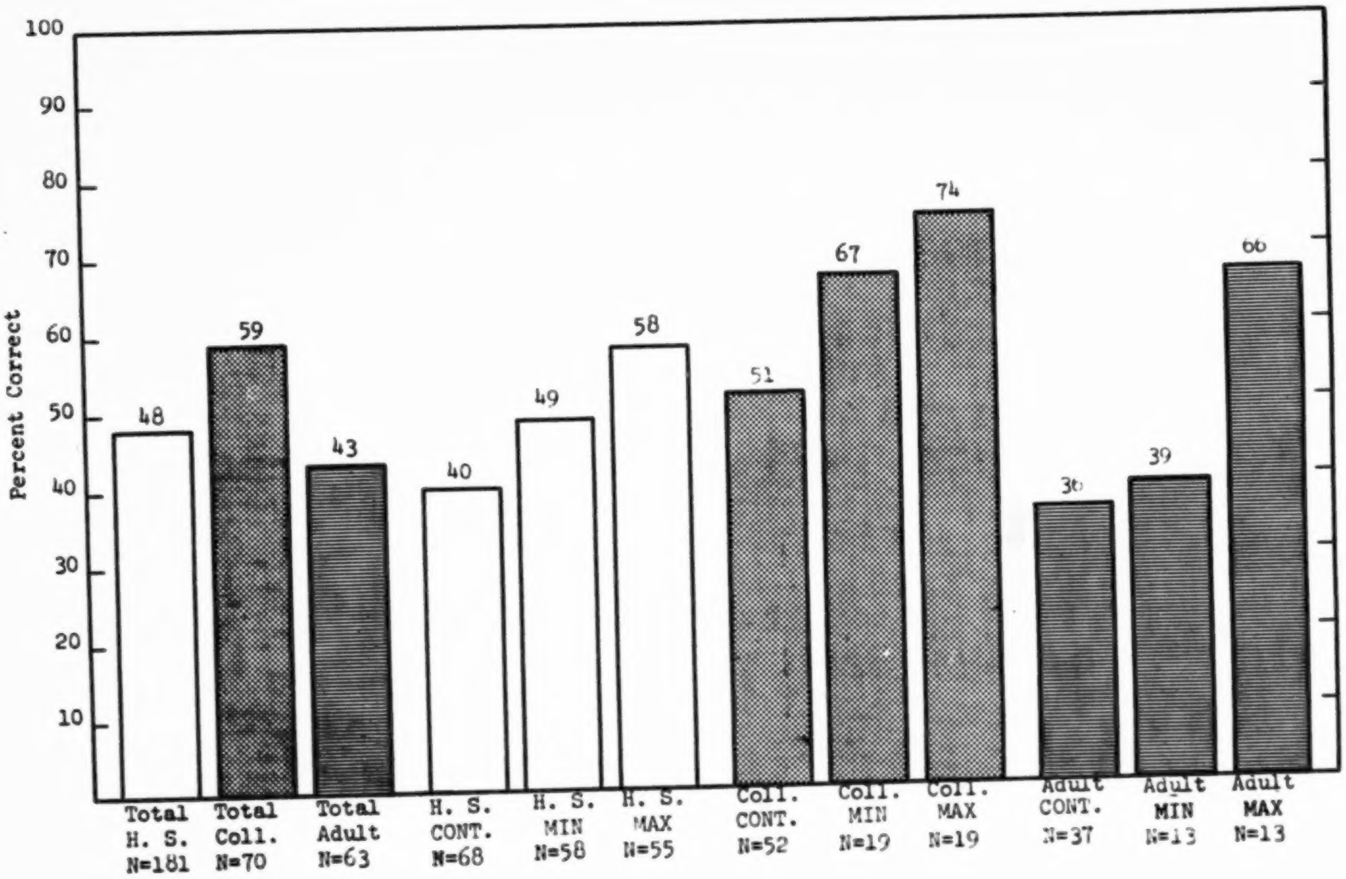


Figure 18. Summary of total knowledge percent correct experimental viewing conditions within age/education groupings.

Results of Analyses of Variance on Selected Variables

This section of the report describes the results of a number of analyses of variance dealing with the following dependent and independent variables. These variables have been described in detail earlier in this report.

Dependent Variables	Independent Variables
1. Interest	1. Viewing conditions
2. Open-end knowledge	2. Age/education
3. Open-end concept	3. Sex
4. Multiple-choice	4. Science background
5. Exhibit-only	5. Socioeconomic status

This more sophisticated method of data analysis was applied primarily as a back-up to the other analyses performed on the data.

The results of these analyses are summarized in Tables 15 through 20. (Individual cell frequencies and means are presented in Appendix M.) Each table summarizes the results of five analyses of variance, one for each of the five dependent measures. Analyses of variance carried out on the attitude data were "garbaged" by the computer and therefore do not appear in this section of the report. Each table, with the exception of Table 20, deals with a different independent variable. The first four tables represent analyses carried out on the same group of subjects, subjects viewing or not viewing the exhibit in Chicago under controlled conditions. These four tables summarize the effects of viewing conditions, age/education, sex, and science background. Table 19, which summarizes the effects of socioeconomic status, was based on a subsample of the total Chicago group. Socioeconomic data was obtained only for the high school subjects. Therefore, the analyses of socioeconomic status could be carried out upon this group only. Table 20 represents a combination of Table 15 with data from the casual viewers of the Chicago exhibit. The data summarized in Table 20 are of primary importance to this study, since a major objective of the study has been to place the casual viewer somewhere along a continuum having meaningful reference points.

The reader who is familiar with factorial designs in analysis of variance may well wonder why the analyses summarized in Tables 15 through 18 were not combined into four-way analyses of variance with a resulting reduction in unaccounted for variance. An examination of the 36 cell frequencies in the four-way classification scheme revealed marked departures from equality. Some of the cells had N's as small as one, and some had N's of 20 or more. Factorial analyses of variance on the obtained data were ruled out because of the likelihood of severe departures from the homogeneity of variance assumptions required. Several factorial analyses of variance were carried out on a trial basis but the wide variability in the within cell variances negated their interpretation. All analyses of variance presented in this section of the report are based on one-way

TABLE 15

Effects of Experimental Conditions
(Based on Experimental Subjects)

DEPENDENT VARIABLE	OBTAINED F VALUE*	MEANS			REQUIRED DIFFERENCE (.05 level)	SIGNIF. COMPARISON		
		1. MIN	2. MAX	3. CONTROL		1-2	1-3	2-3
Interest	23.95	16.48	17.00	14.97	.61		X	X
O-E Concept	39.58	9.13	12.36	7.03	1.52	X	X	X
O-E Knowledge	51.07	15.97	19.93	10.46	2.43	X	X	X
Multiple Choice	20.01	45.51	53.63	42.29	4.64	X		X
Exhibit Only	54.99	11.57	13.45	7.93	1.32	X	X	X

*F needed for significance at .05 level with 2 and 331 degrees of freedom is 3.03.

TABLE 16

Age/Education Differences
(Based on Experimental Subjects)

DEPENDENT VARIABLE	OBTAINED F VALUE*	MEANS			REQUIRED DIFFERENCE (.05 level)	SIGNIF. COMPARISON		
		1. H.S.	2. COLL.	3. ADULT		1-2	1-3	2-3
Interest	3.34	16.18	15.74	15.33	.69		X	
O-E Concept	13.59	8.44	11.29	7.25	1.75	X		X
O-E Knowledge	7.63	14.26	17.09	10.98	2.88		X	X
Multiple Choice	16.84	43.83	54.17	41.16	4.96	X		X
Exhibit Only	8.40	10.29	11.86	8.37	1.57	X	X	X

*F needed for significance at .05 level with 2 and 331 degrees of freedom is 3.03.

TABLE 17

Sex Differences
(Based on Experimental Subjects)

DEPENDENT VARIABLE	OBTAINED F VALUE*	MEANS		SIGNIFICANCE OF COMPARISON
		MALE	FEMALE	
Interest	.11	16.01	15.79	n.s.
O-E Concept	4.87	10.01	7.79	X
O-E Knowledge	8.12	16.22	12.28	X
Multiple Choice	10.72	49.88	41.70	X
Exhibit Only	2.24	11.17	9.38	n.s.

*F needed for significance at .05 level with 1 and 332 degrees of freedom is 3.87.

TABLE 18

Science Background Differences
(Based on Experimental Subjects)

DEPENDENT VARIABLE	OBTAINED F VALUE*	MEANS		SIGNIFICANCE OF COMPARISON
		SCIENCE	NONSCIENCE	
Interest	1.06	16.11	15.70	n.s.
O-E Concept	11.45	10.20	7.75	X
O-E Knowledge	7.27	16.20	12.57	X
Multiple Choice	13.14	50.15	41.98	X
Exhibit Only	3.16	11.34	9.33	n.s.

*F needed for significance at .05 level with 1 and 332 degrees of freedom is 3.87.

TABLE 19

Effects of Socioeconomic Status
(Based on Experimental High School Subjects)

DEPENDENT VARIABLE	OBTAINED F VALUE*	MEANS			REQUIRED DIFFERENCE	SIGNIF. COMPARISON		
		1. HIGH	2. MEDIUM	3. LOW		1-2	1-3	2-3
Interest	.22	15.88	16.37	16.31	----			
O-E Concept	4.59	8.59	9.77	7.80	2.10	**		
O-E Knowledge	3.13	14.46	16.63	13.17	3.34			X
Multiple Choice	11.10	44.69	50.77	40.45	5.97	X		X
Exhibit Only	4.85	10.20	12.17	9.59	2.01			X

*F needed for significance at .05 level with 2 and 178 degrees of freedom is 3.05.

**Although a significant F value was obtained, none of the comparisons of individual treatment means are significant. Other comparisons involving combinations of means, not examined here, account for the significant F.

TABLE 20

A Comparison of Casual Viewers with Experimental Viewers and Nonviewers
(Based on All Subjects)

DEPENDENT VARIABLE	OBTAINED F VALUE*	MEANS				REQUIRED DIFFERENCE	SIGNIFICANT COMPARISONS					
		1. CASUAL	2. MIN	3. MAX	4. CONTROL		1-2	1-3	1-4	2-3	2-4	3-4
Interest	14.88	15.08	16.48	17.00	14.97	1.02	X	X			X	X
O-E Concept	38.56	4.65	9.13	12.36	7.03	2.56	X	X		X		X
O-E Knowledge	47.34	6.38	15.97	19.93	10.46	4.19	X	X			X	X
Multiple Choice	39.89	28.77	45.51	53.63	42.29	7.88	X	X	X	X		X
Exhibit Only	70.16	4.58	11.57	13.45	7.93	2.29	X	X	X		X	X

*F needed for significance at .05 level with 3 and 356 degrees of freedom is 2.63.

classification schemes. This was done in order to meet the required assumptions regarding homogeneity of variance. Examination of the calculated variances within each of the treatments in the one-way classification schemes revealed no serious departures from the assumption of equal within treatment variances.

An examination of Table 15 reveals that the viewing conditions produced, for the most part, the desired effects. The maximum exposure condition generally produced more favorable results than the minimum exposure condition, while the minimum exposure condition generally produced more favorable results than the control or non-viewing condition. There are two exceptions to this general statement. On interest measures, the MIN and MAX groups were not significantly different from each other. The second exception occurred with respect to the multiple-choice test where the minimum exposure group did not differ significantly from the CONTROL group. Although these two comparisons yielded nonsignificant differences it should be pointed out that the direction of the differences that were obtained is consistent with the overall trend of the data. The five analyses summarized in Table 15 have accomplished two purposes:

1. They have demonstrated that the five instruments are able to discriminate among the effects of various exposure times for highly motivated subjects; and
2. They have provided a series of bench marks or reference points with which the behavior of casual viewers may be compared.

The analyses summarized in Tables 16 through 19 deal with viewer characteristics, frequently referred to as organismic variables since they represent a property of the organism that is not ordinarily subject to experimental manipulation or control. One cannot randomly assign age or sex to subjects in the same fashion as assigning subjects to various experimental conditions!

Two general patterns tend to occur in the data presented in Table 16 summarizing the effects of the age/education variable, although quite a few of the individual comparisons were not significant. High school students had significantly more interest than adults, with college students falling practically half way between the two and not differing significantly from either group. This pattern suggests that age, rather than education, is the distinguishing factor, since the educational level of the adults sampled was lower than that of the college students. Education, rather than age, would appear to account for the pattern obtained on the four remaining measures. Without exception, on these measures, the college students were superior to the adults. The college students performed significantly better than adults on all four measures and better than high school students on three of the four.

The sex differences summarized in Table 17 are not surprising in view of the fact that males are more likely to have had past training in science. Although the differences in interests for the two sexes are in the expected direction, they are not significant. The same can be said for the exhibit-only measure. A comparison of Table 18 with Table 17 shows very similar patterns of means suggesting that the sex differences may be largely explained in terms of the differing science backgrounds for males and females.

The effects of socioeconomic status are summarized in Table 19. Recall that these data are based on high school students only, as mentioned previously. No significant effects of socioeconomic status upon interest were observed. The general trend for the remaining four measures was for subjects from the middle socioeconomic level to perform better than those from the lower level, while subjects from the upper level performed at a level between these two groups. This may in part be a function of the achievement orientation of the middle classes with respect to education and technology. However, additional data analyses would be required, using more detailed breakdowns of IQ, reading level, etc., to provide substantive information on the reason for this trend in the data.

A comparison of the casual viewer with respect to the three experimental conditions is summarized in Table 20. The reader will note that some of the comparisons involving the MIN, MAX, and CONTROL conditions failed to attain significance here although they were described as significant in Table 15. For example, the CONTROL group has previously been described as being significantly different from the MIN group with respect to the open-end concept measure. This comparison, however, is reported as not significant in Table 20, (although it does come within .46 of meeting the requirements for significance). These apparent inconsistencies reflect the fact that different estimates of error-variance are being used in the two analyses. Casual viewers do not differ significantly from the CONTROL group with respect to interest, and are significantly lower than both the MIN and MAX groups along the interest dimension.

The general trend for the remaining four knowledge measures is for the casual viewers to be below the other three groups in terms of test performance. They do not differ significantly from the CONTROL groups on the two open-end measures but are significantly poorer than the CONTROL group with respect to the multiple-choice and exhibit-only tests. Reasons for this anomaly have been discussed earlier. The fact that it appears as a significant factor in the analysis of variance data lends additional support to its reality, whatever may be the causes of its existence.

Figures 19 through 23 contain idealized distributions based on the obtained means and variances for the casual viewers and experimental subjects for each of the five measures. A striking feature of these distributions is the large amount of variance within and overlap between the various groups. If the sample populations

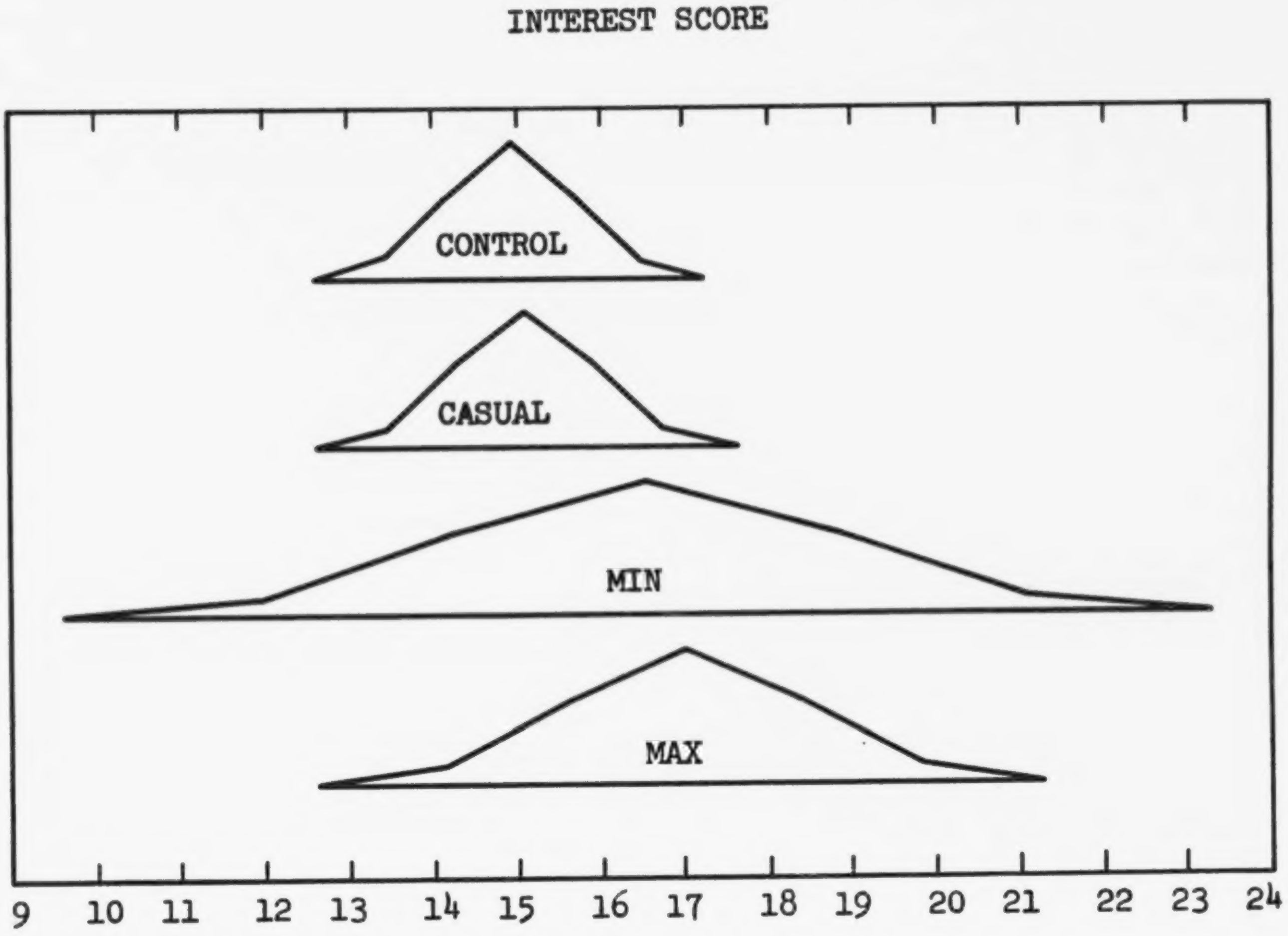


Figure 19. Idealized distributions of interest scores for the four groups.

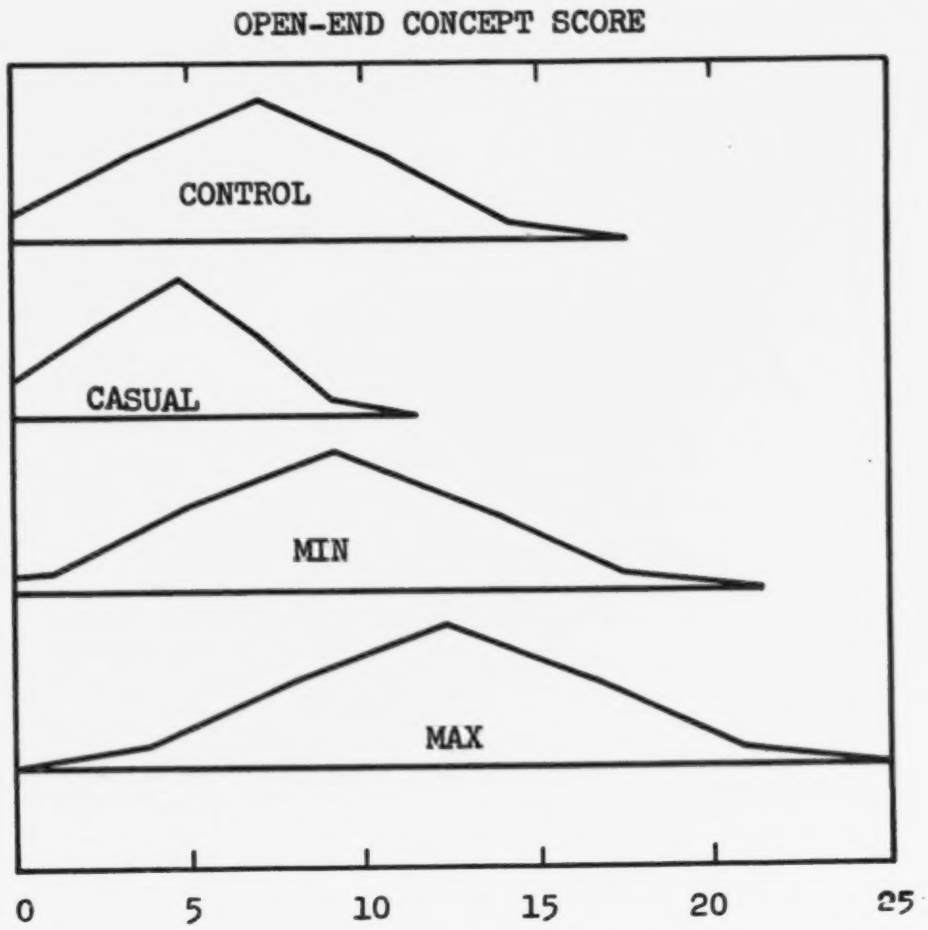


Figure 20. Idealized distributions of open-end concept scores for the four groups.

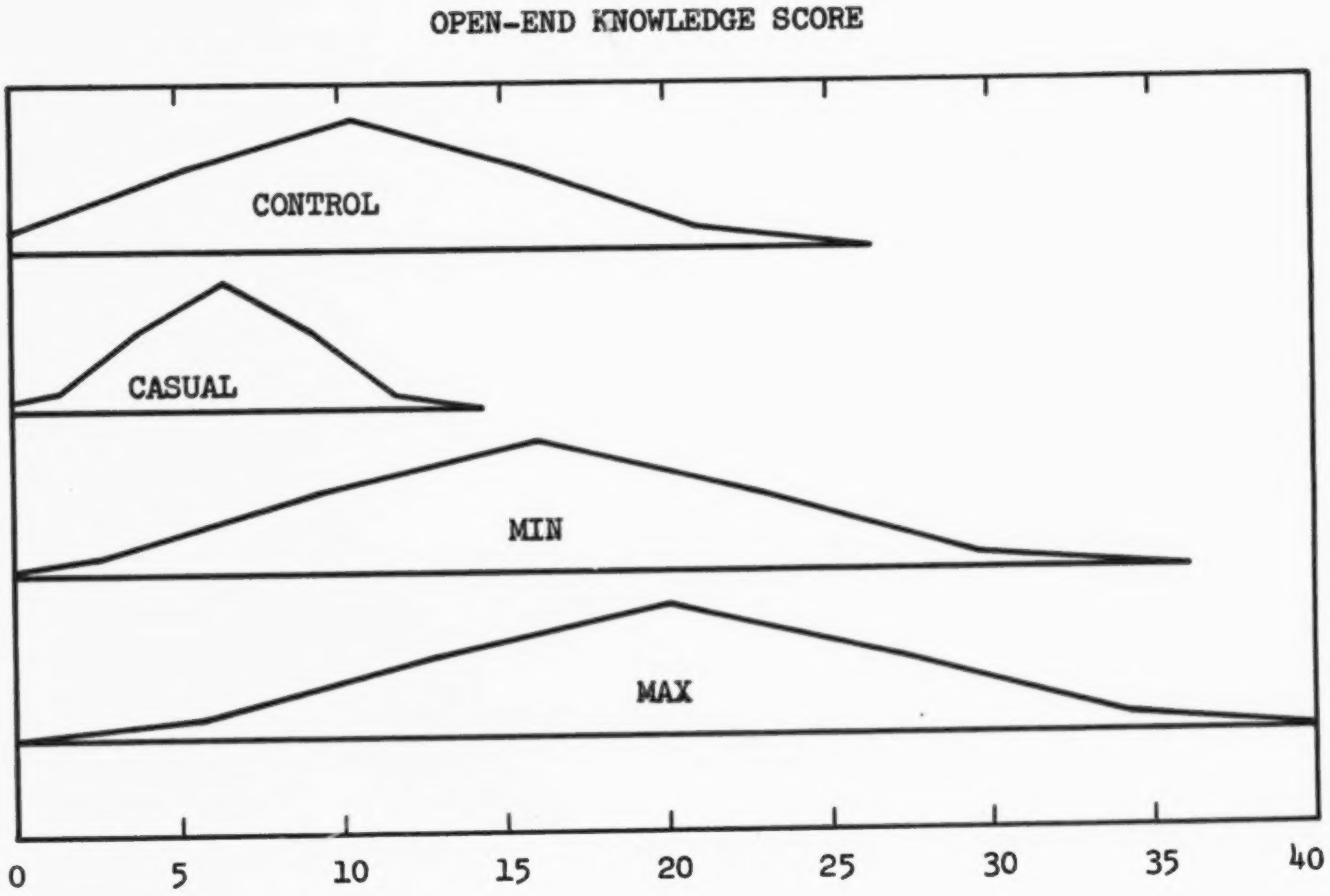


Figure 21. Idealized distributions of open-end knowledge scores for the four groups.

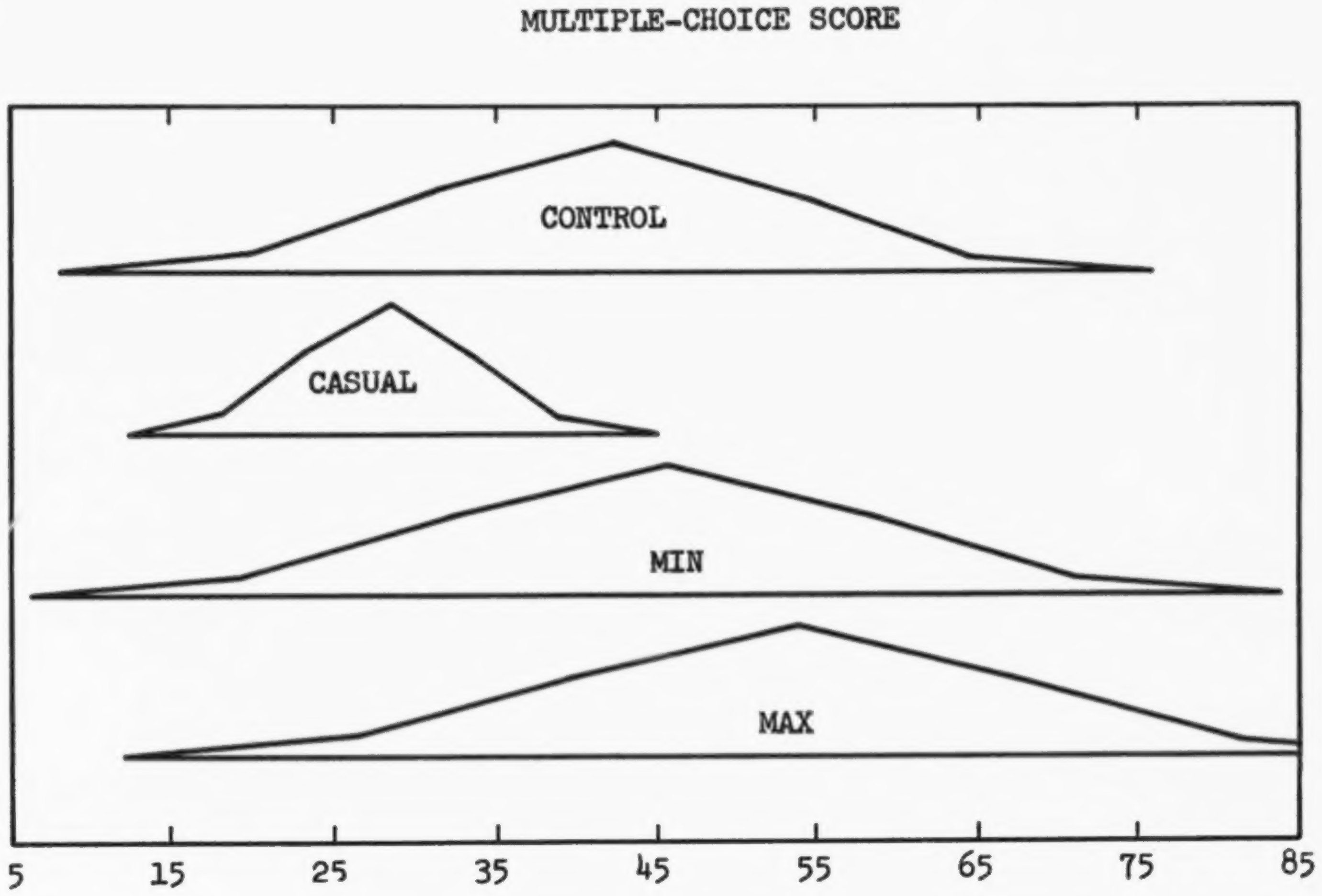


Figure 22. Idealized distributions of multiple-choice scores for the four groups.

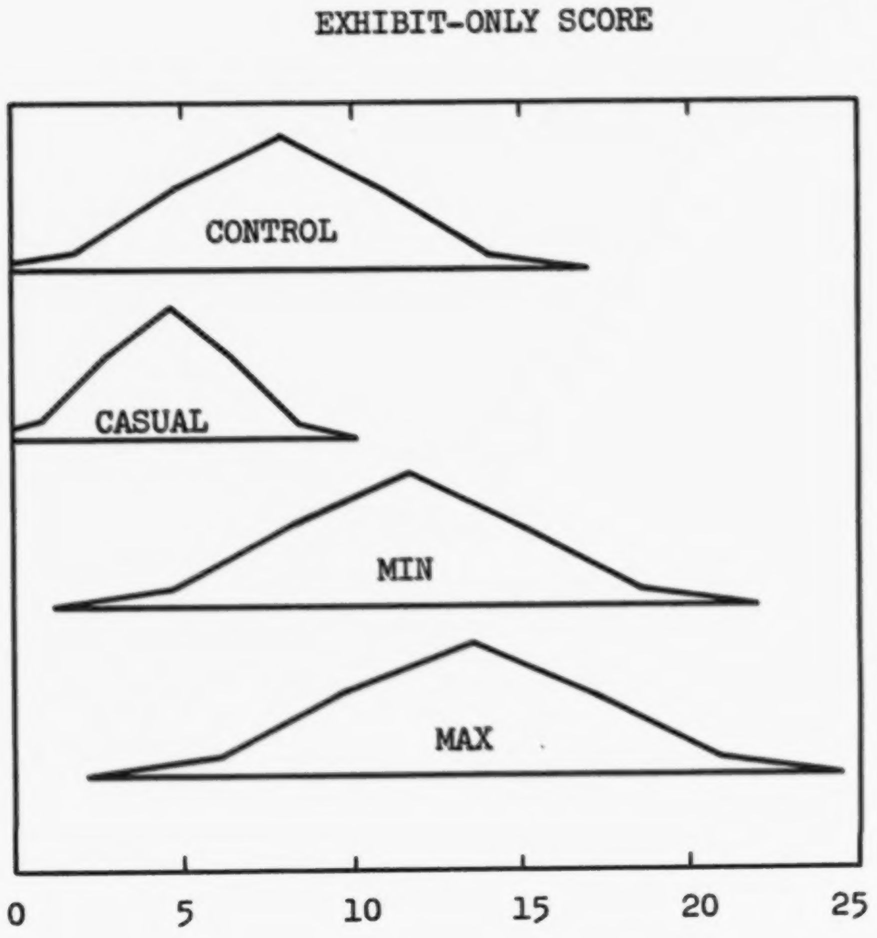


Figure 23. Idealized distributions of the exhibit-only scores for the four groups.

used in this study were representative of the true populations, and if the tests used were reliable and valid measures of the various areas, then the distribution and means shown are those that actually exist in the various categories of exhibit viewers in the real world. The variance measures obtained from casual viewers is, without exception, less than that obtained on subjects in the other conditions. Several factors may account for this phenomenon, although no definite conclusions may actually be drawn. One possibility centers around the small number of cases in the casual viewing condition. This alone could produce lower variances. A second possibility is that a "floor effect" reduced variability, i.e., the casual viewers couldn't get much worse! A third possibility centers around the fashion in which the casual viewers' scores were "constructed" from the responses of a greater number of actual casual viewers. As mentioned elsewhere in this report, data from sets of subjects who completed only portions of the actual measuring instruments were combined to produce complete data on "psuedo-subjects." These combinations of data thus may be considered as sample means based on a number of cases. According to the central limit theorem, the variation in means of samples may be expected to be much less than the actual variation of individual cases in the population. Although it cannot be stated that this latter explanation is sufficient to account for the phenomenon, it appears to be at least partially responsible for the reduced variability in the casual viewer scores.

Mock-up Validation Results

The basic approach to validating the mock-up consisted of a comparison of systematic variations across various features of the exhibit with corresponding variations across features of the mock-up and of a comparison of the absolute level of learning. The former analysis will be discussed first.

If the mock-up is a valid representation of the exhibit, then the order in which subareas of the actual exhibit may be rank ordered along some criterion should be essentially the same as the rank order of the corresponding subareas of the exhibit mock-up along the same criterion. This approach, of course, ignores the fact that the actual differences between a subarea of the exhibit and the corresponding subarea of the mock-up may be quite great. (This approach is discussed later.) A valid mock-up of an exhibit may do a poor job of replicating the actual levels of the criterion values that one may obtain from a real exhibit and still rank order the subareas of the exhibit quite accurately.

Data used in the initial mock-up validation was based on subjects in the MIN and MAX conditions in the actual exhibit at Chicago and in the MIN and MAX conditions at the mock-up in Pittsburgh. Questions from each of the six instruments were related to specific areas, or subareas, of the exhibit.

In case of the attitude measures, the specific items were related to the five main objectives of the exhibit. Each objective had associated with it a specific question or set of questions on the attitude instrument designed to measure the attainment of that objective. Thus, the five objectives of the exhibit could be arranged in rank order by placing the objective that was most fully attained by the exhibit at the top of the list, the objective that was next to that one in degree of attainment immediately below it, and so forth. This procedure was carried out for the five objectives for both the mock-up and the exhibit group. The resulting five pairs of ranks were correlated using Spearman's rank order correlation method. The obtained correlation, designated as Rho in Table 21, was .90. This correlation was significant at the .05 level of confidence. Thus, the two procedures, the mock-up and the actual exhibit, produced highly similar rank ordering of the objectives in terms of their successful attainment. (A correlation of .90 means that 81 percent of the variance in the attitude rankings of objectives by the mock-up was common to the variance in the rankings produced by the actual exhibit.)

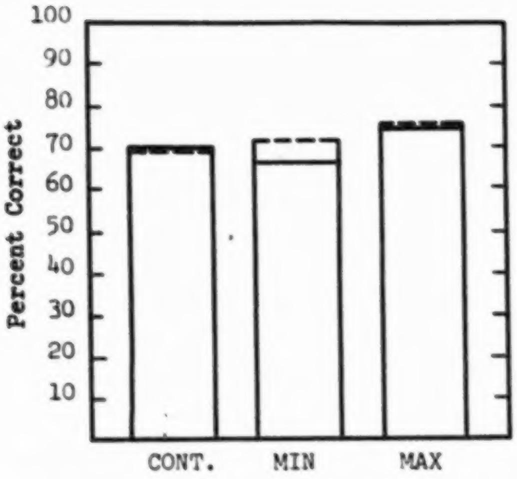
TABLE 21

Mock-up Validation Results
(Based on Experimental and Mock-up Subjects)

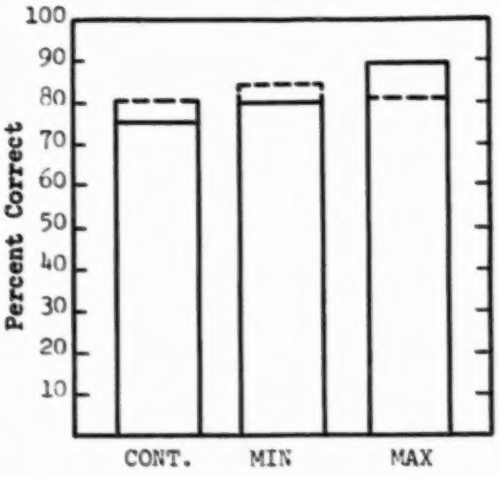
Dependent Variable	Correlated across	Rho	Signif.*	Rho′	Signif.*
INTEREST	5 areas	.60	n.s.	.60	n.s.
ATTITUDE	5 objectives	.90	.05	.23	n.s.
O-E CONCEPT	14 subareas	.92	.01	.64	.05
O-E KNOWLEDGE	13 subareas	.70	.05	.89	.01
MULTIPLE CHOICE	26 subareas	.87	.01	.80	.01
EXHIBIT ONLY	11 subareas	.57	.05	.14	n.s.

*Based on levels at which rho is significant using a one tailed test.

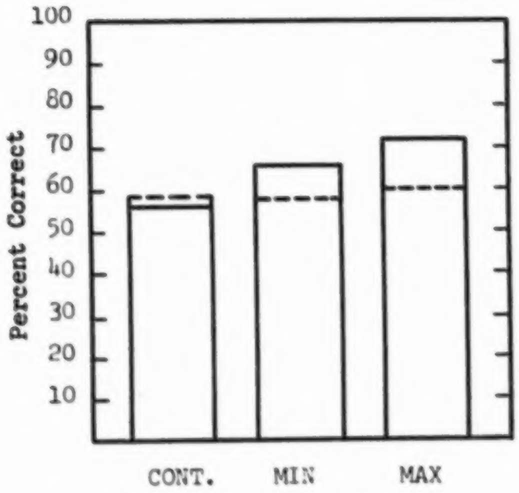
Figure 24 shows the mean attitude scores of the mock-up experimental groups for each exhibit objective. The dotted lines indicate the mean attitude scores for the Chicago high school subjects. Although there are some differences between the mock-up and exhibit scores on each of the objectives, these differences are not significant. Notice that in some cases, the exhibit group attained a higher mean score while in other cases, the mock-up group attained a higher score.



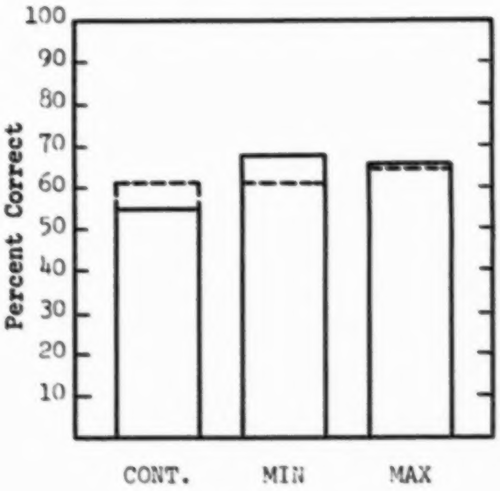
Objective #1
Basic achievements and
impact of science.



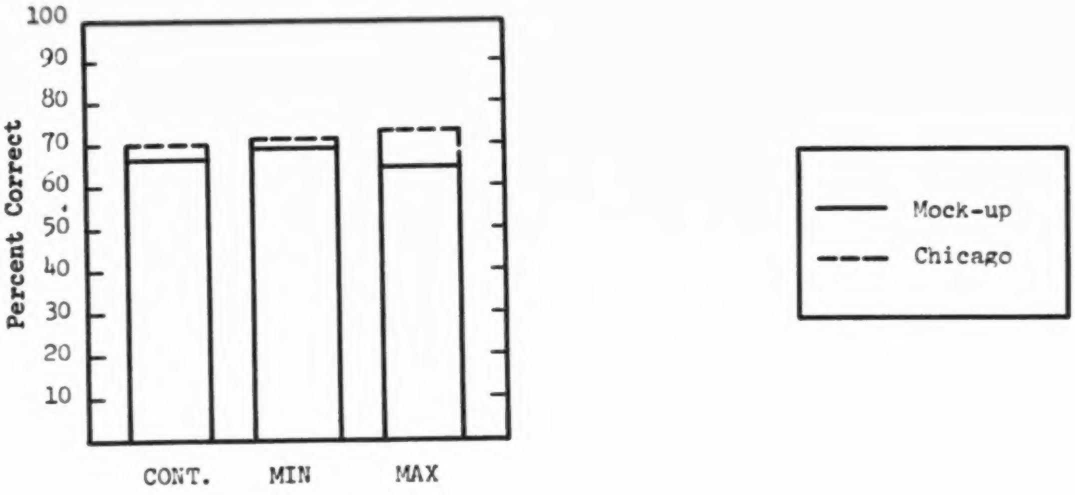
Objective #2
Awareness of the interaction
between different sciences
and between science and technology.



Objective #3
Awareness of (and regard for)
relation between Federal
Government and scientific projects.



Objective #4
Regard for scientists and
scientific projects.



Objective #5
Interest in studying science
and in a scientific career.

Figure 24. Mock-up and Chicago mean attitude scores for each exhibit objective.

Within each exhibit objective, the similarity between the control group scores and the MIN and MAX scores indicates that basic attitudes are not changed significantly by viewing either an exhibit mock-up or the actual exhibit.

Procedures similar to the above were carried out for the interest measures, and the remaining four knowledge measures. The remaining five measures were correlated across either areas or subareas of the exhibit, rather than objectives. The resulting correlations are summarized in the column labeled Rho in Table 21. All of the correlations are significant, with the exception of the one based on interest measures. Figure 25 shows the mock-up mean interest ranks for each exhibit area. When this figure is compared with the Chicago high school group (Figure 9 on page 55), it is apparent that the mock-up and Chicago subjects ranked the exhibit areas differently in terms of interest. The low correlation obtained in the comparison between mock-up and exhibit interest scores is thus substantiated by these data. When the mock-up CONTROL group interest pattern is compared with the interest patterns of the other two experimental groups, it is apparent that interest levels are changed by viewing an exhibit mock-up.

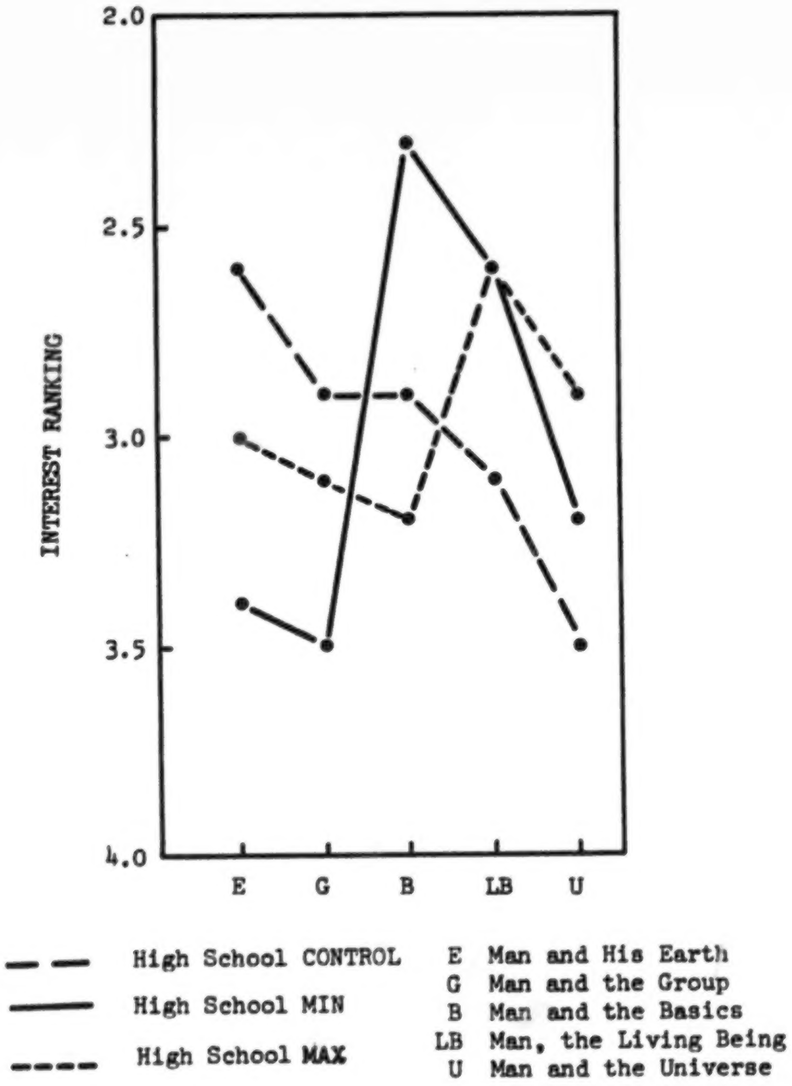


Figure 25. Mean interest ranks for MIN, MAX, and CONTROL high school subjects in mock-up validation.

The mock-up attitude and interest results support the conclusions based on the exhibit testing phase. It appears that interest levels can be changed by viewing either an exhibit mock-up or the actual exhibit, although the direction and extent of the changes were not comparable. The exhibit attitude data which are replicated in the mock-up results support the conclusion that there is a general leveling phenomenon in attitude scores independent of exhibit stimuli.

All of the correlations between the mock-up and the exhibit seem rather high, suggesting that the mock-up did an exceptionally good job of replicating the actual exhibit. It is quite possible, however, that these high correlations arose primarily from variations in item difficulty for the various knowledge measures or from stereotypical responding to the interest and attitude measures. In order to investigate these possibilities, it was necessary to get baseline data on difficulty levels of the questions dealing with the various subareas and areas of the exhibit and also to measure the typical responses to the interest and attitude items among subjects not viewing the exhibit. Control subjects from the designated Chicago CONTROL group were used to provide baseline data for all six measures. Instead of correlating ranks based on mean scores for subareas, as done previously, the mean scores were corrected with respect to the control condition. Thus, if the percentages of correct responses to a particular subarea averaged 65 percent for the actual exhibit, 30 percent for the CONTROL group and 50 percent for the mock-up, the percentage for the CONTROL group was subtracted from both the exhibit and mock-up groups, resulting in "gain scores" of 35 percent and 20 percent for the two groups on the particular subarea considered. This procedure produced sets of gain scores on each of the six measures for each of the relevant features of the exhibit, i.e., the subareas, areas, and objectives. The gain scores were separately rank ordered for the exhibit and mock-up conditions and then correlated using the same procedure described earlier. The obtained correlations are presented in the column labeled Rho′ in Table 21 and represent correlations between the effectiveness of the relevant features of the exhibit and the mock-up with the effects of the difficulty and stereotypical responding partialed out. With one exception (open-end knowledge) these correlations are lower than those previously obtained, indicating that item difficulty and stereotypical responding accounted for a portion of the apparent "validity" of the mock-up.

Figures 26 and 27 are scatterplots of the mean open-end concept scores on each of 14 exhibit subareas. Figure 26 shows the initially high correlation of .92. Figure 27 shows the second correlation of .64, corrected for the effects of variation in item difficulty. This represents a considerable reduction in variance common to the two conditions, i.e., from 85 percent (.92 squared) to 41 percent (.64 squared). Thus, more than half of the variance common to the mock-up and the actual exhibit was accounted for by an "irrelevant" factor, item difficulty, common to both sets of

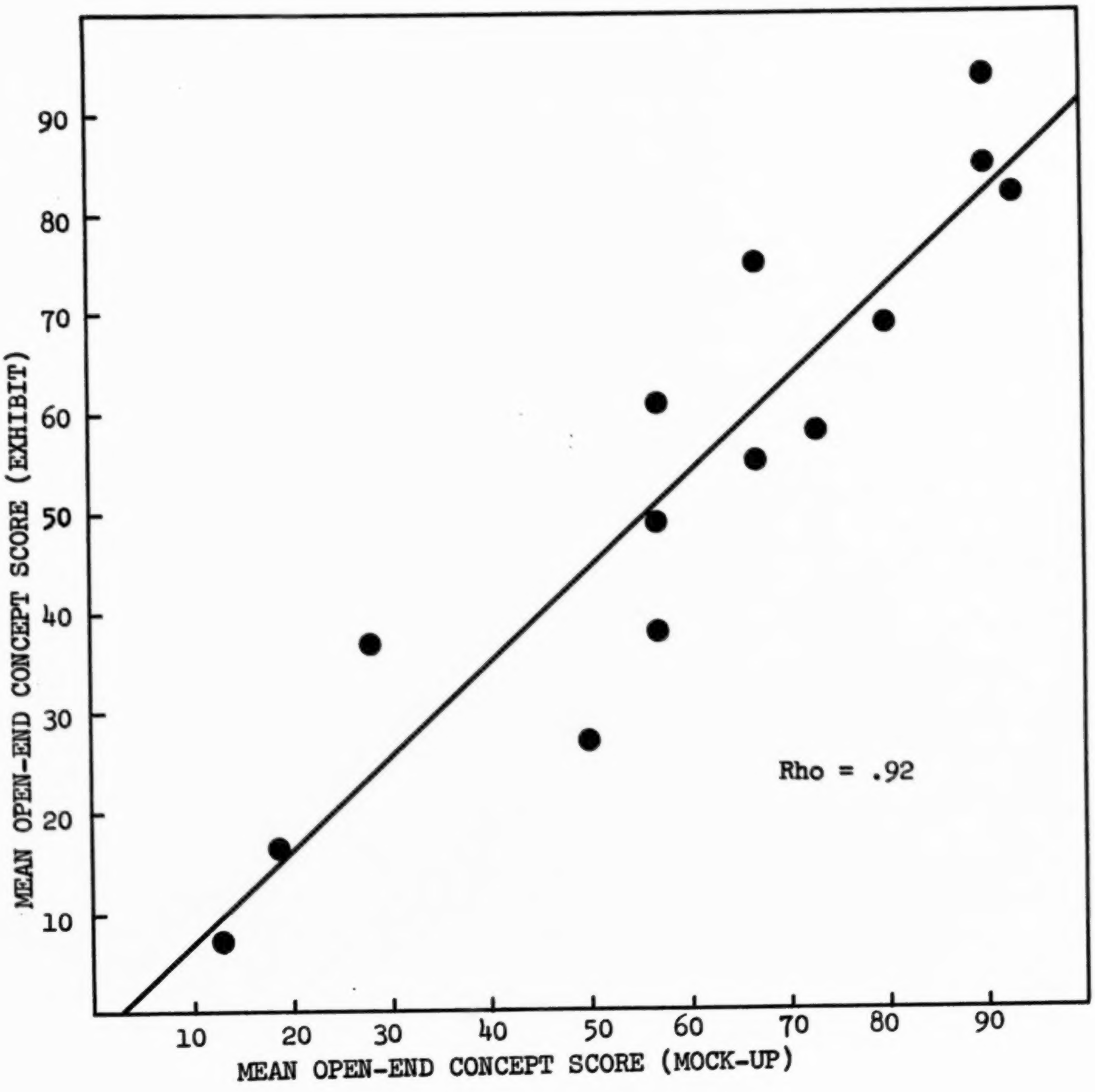


Figure 26. Relationship between actual exhibit and exhibit mock-up in terms of uncorrected mean scores for 14 subareas.

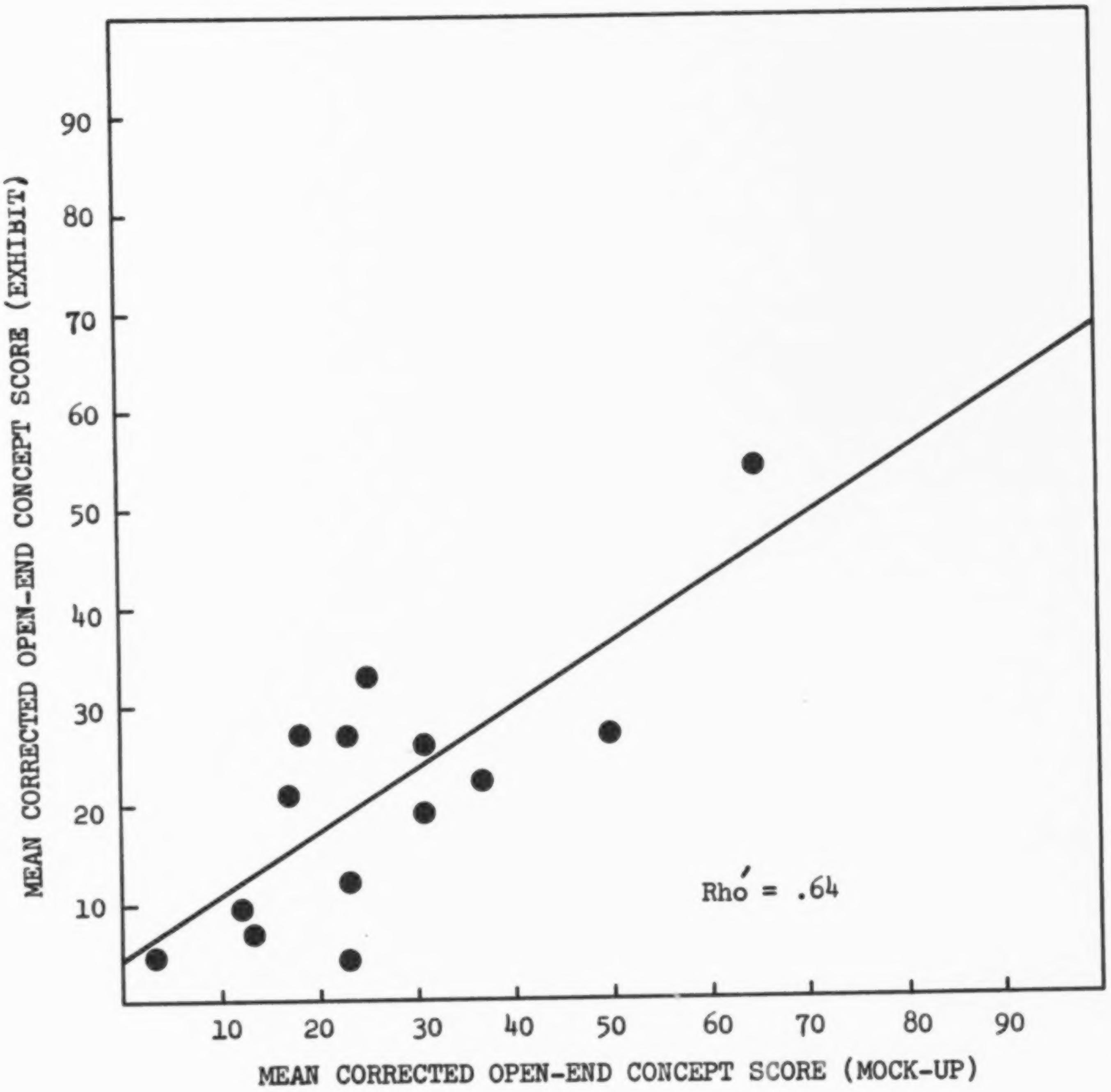


Figure 27. Relationship between actual exhibit and exhibit mock-up in terms of corrected mean scores for 14 subareas.

measures. This phenomenon is repeated with the attitude and exhibit-only measures. The interest, open-end knowledge, and multiple-choice measures do not "shrink" as much when the correction is made. In fact, the open-end knowledge results are improved by the corrections.

It is felt that the column labeled Rho′ in Table 21 gives the more accurate picture of the "validity" of the mock-up. Relative effectiveness indices for the various subareas of the exhibit on three of the four knowledge measures are fairly predictable from the mock-up data. The interest, attitude, and exhibit-only measures for the areas, objectives, and subareas respectively, are not predictable from the mock-up data. This is not particularly surprising in view of the small N (5) for the interest and attitude measures. For an N of 5, Spearman's Rho must be .90 or greater to be significant at the .05 level. Thus, although the interest and attitude measures may demonstrate fair relationships between the exhibit and the mock-up, these relationships do not meet customary significance levels. These results point toward the necessity for a more detailed breakdown of features of the exhibit and mock-up in order to determine their validity. They also suggest that certain effects of an exhibit, particularly in the affective as opposed to the cognitive domain, may be more difficult to capture within a mock-up.

Comparisons of the actual levels of performance of subjects viewing the exhibit were also made with subjects viewing the mock-up for each of the six measures. These analyses were carried out separately for the MIN and MAX groups, resulting in a total of 12 comparisons. Note that these analyses were carried out without regard to subareas. If the mock-up and the exhibit were comparable in actual teaching effectiveness, the difference between the scores should _not_ be significant.

With one exception, all 12 comparisons yielded nonsignificant differences between the exhibit and mock-up groups. The one analysis yielding significant results compared the performance of mock-up and exhibit subjects on the multiple-choice test under minimum exposure conditions. Mock-up subjects scored significantly _higher_ than exhibit subjects ($t = 2.43$, 71 d.f., $p < .02$). Although this finding is somewhat surprising, it may indicate that the mock-up contained fewer distractors (interesting but relatively uninformative displays) than the exhibit. Thus, the subject pressed for time under the minimum exposure condition would be penalized most by such distractors. Performance of the exhibit and mock-up subjects under the _maximum_ exposure condition was almost identical ($t = .330$, 68 d.f.), suggesting that the reduced amount of viewing time may have been responsible for the inferior performance of subjects viewing the actual exhibit under minimum exposure conditions.

In general, the overall results of the mock-up validation are most encouraging. They demonstrate that useful information about the relative effectiveness of various features of the exhibit may be determined with a fair degree of accuracy. In addition, it was demonstrated that the actual levels of knowledge, attitude, and interest that occurred among subjects viewing the exhibit could be predicted on the basis of mock-up data. In summary, the results of the mock-up validation study show that the construction of mock-ups or small-scale simulations of exhibits can provide the exhibit designer with relatively inexpensive and useful information regarding the future performance of this exhibit. The use of such mock-ups also permits a much higher degree of control and manipulation of exhibit design variables. As has been noted, the use of a mock-up strategy closely parallels the pretesting that goes into the construction of modern educational materials before a "hands off" trial takes place. In view of the costs of exhibits such as the Vision of Man, the costs of constructing small-scale mock-ups for pretesting purposes would appear to be quite small.

Mock-up Variation Results

The basic percentage scores for the eight experimental groups are shown in Tables 22 and 23. In Table 22, the scores are broken down according to test. In Table 23, they are broken down according to subject matter. Group 6, Viewers -- Read Skeleton Text, did better than all other groups both in total knowledge and in the breakdowns, except on the open-end concept questionnaire. Group 8, Viewers -- Heard Full Text, did worse than all other groups except in the open-end knowledge questionnaire.

Table 24 shows the total knowledge scores for each group placed in their category under each of the three experimental variables. An analysis of variance of these three variables (dependent variable = total knowledge scores) indicated no significant difference between viewing or not viewing the mock-up, or between skeleton text and full text. The difference between reading the text and hearing it is significant at the .001 level of confidence. No significant interaction of the variables was found. Table 25 summarizes the results of the analysis of variance.

IQ scores were obtained for all subjects participating in the mock-up variation study. Table 26 shows the average IQ scores for each group, along with their within-cell correlation with total knowledge scores. Only one group -- Group 4, Non-viewers, Heard Skeleton Text -- shows a significant within-cell correlation between IQ and total knowledge.

The overall correlation between IQ and total knowledge, however, across all conditions, is +.43 and is highly significant. Since the subject population was made up of very intelligent students (Mean IQ = 126, standard deviation = 8.44) the range of the

TABLE 22

Mock-up Variation Percentage Scores by Group and by Test

	Open-End Concept	Test Open-End Knowledge	Multiple Choice	Exhibit Only	TOTAL
Nonviewers; Read Full	67%	65%	78%	74%	74%
Nonviewers; Read Skeleton	60%	68%	74%	67%	70%
Nonviewers; Heard Full	47%	49%	68%	68%	61%
Nonviewers; Heard Skeleton	57%	65%	66%	71%	66%
Viewers; Read Full	67%	66%	76%	83%	74%
Viewers; Read Skeleton	63%	82%	89%	86%	83%
Viewers; Heard Skeleton	52%	65%	66%	71%	65%
Viewers; Heard Full	43%	51%	59%	64%	56%
Average	57%	64%	72%	73%	69%

TABLE 23

Mock-up Variation Percentage Scores by Group and by Subject Matter

	Amplification, Extension, and Substitution	Fish Sounds	Chrysanthemum	Cosmic Rays	TOTAL
Nonviewers; Read Full	77%	80%	73%	70%	74%
Nonviewers, Read Skeleton	70%	83%	74%	63%	70%
Nonviewers; Heard Full	63%	65%	66%	56%	61%
Nonviewers; Heard Skeleton	62%	58%	64%	71%	66%
Viewers; Read Full	77%	78%	71%	74%	74%
Viewers; Read Skeleton	85%	90%	86%	79%	83%
Viewers; Heard Skeleton	67%	63%	69%	62%	65%
Viewers; Heard Full	55%	57%	62%	52%	56%
Average	69%	72%	70%	66%	69%

TABLE 24

Mock-up Variation Total Knowledge Percentage Scores According to the Three Variables of the Study

Variable A: Viewers vs Nonviewers

Nonviewers (N=40)		Viewers (N=40)	
Group	TK Score (%)	Group	TK Score (%)
Read Full	74	Read Full	74
Read Skeleton	70	Read Skeleton	83
Heard Full	61	Heard Skeleton	65
Heard Skeleton	66	Heard Full	56
Total	68	Total	70

Variable B: Sound vs Written

Heard (N=40)		Read (N=40)	
Group	TK Score (%)	Group	TK Score (%)
Nonview; Full	61	Nonview; Full	74
Nonview; Skeleton	66	Nonview; Skekelton	70
View; Skeleton	65	View; Full	74
View; Full	56	View; Skeleton	83
Total	62	Total	75

Variable C: Full text vs Skeleton Text

Full (N=40)		Skeleton (N=40)	
Group	TK Score (%)	Group	TK Score (%)
Nonview; Read	74	Nonview; Read	70
Nonview; Heard	61	Nonview; Heard	66
View; Read	74	View; Read	83
View; Heard	56	View; Heard	65
Total	66	Total	71

TABLE 25

Analysis of Variance of Three Mock-up Variation Variables (Dependent Variable = Total Knowledge Scores)

Variable A = Visuals (Viewed vs Didn't View)

Variable B = Media (Heard vs Read)

Variable C = Text (Full vs Skeleton)

Dependent Variable = Total Knowledge Scores

Variable	Degrees of Freedom	Sum of Squares	Mean Square	F Ratio	p
A	1	13.61	13.61	0.357	n.s.
B	1	621.61	621.61	16.281	.001
C	1	86.11	86.11	2.255	n.s.
AxB	1	86.11	86.11	2.255	n.s.
AxC	1	63.01	63.01	1.650	n.s.
BxC	1	15.31	15.31	0.401	n.s.
AxBxC	1	21.01	21.01	0.550	n.s.
Within	72	2748.90	38.18		
Total	79	3655.69			

TABLE 26

Correlation Between IQ and Total Knowledge for Each Group in Mock-up Variation Study

	Average IQ	Average TK Score (%)	r	Sig
Nonviewers; Read Full	124	74	-.10	--
Nonviewers; Read Skeleton	126	70	+.32	--
Nonviewers; Heard Full	124	61	+.58	--
Nonviewers; Heard Skeleton	124	66	+.88	.01
Viewers; Read Full	130	74	+.25	--
Viewers; Read Skeleton	130	83	+.50	--
Viewers; Heard Skeleton	127	65	+.34	--
Viewers; Heard Full	126	56	+.41	--
Average	126	69	+.44	--

IQ measures was much less than that which would occur in the general population. Therefore, the correlation of +.43 is probably an underestimate of the correlation that would be obtained in the general population. An estimate based on the sample correlation and variances of the probable correlation between IQ and total knowledge scores for the general population is +.88. [The technique for computing this is found in Gulliksen (12).]

As described in the Methodology section, each group contained five science-oriented subjects and five nonscience-oriented subjects. The average score for science subjects (across all groups) is 73 percent; for nonscience subjects, 65 percent. The difference is significant at the .02 level (t = 2.4547). Figure 28 shows the relationship between the scores of science subjects and nonscience subjects by group. Note that, except for Group 7 and 8 (viewers, heard text; Group 7 heard skeleton text, Group 8 heard full text), the same general trend is shown in both the science and nonscience groups. That is, Group 3 is lower than the other five, Group 6 is higher, Group 1 is approximately average. Note that the Group 6 nonscience subjects scored higher than most groups' science subjects. A t test reveals a significant difference between the average IQ of science subjects (129) and the average IQ of nonscience subjects (124). (t = 2.6130, significant at the .02 level of confidence.) Thus, the difference in the performance of science and nonscience subjects may reflect differences in intelligence or scientific training or both.

Mock-up Variation Discussion

The mock-up variations study had two purposes. One was to answer the question, "Once a valid mock-up has been prepared, can it be used to identify the relative teaching effectiveness of the exhibit and the portions that may need to be changed?" The other purpose concerned the validity of using an exhibit mock-up as a useful research tool for studying exhibit design variables. Three common exhibit variables were manipulated using the mock-up:

1. Use of visual effects
2. Level of text presented
3. Manner of text presentation

Use of visual effects. The analysis of variance failed to show a significant difference between those who viewed the mock-up with visuals and those who didn't. In fact, the F ratio of 0.357 is extremely low. Since, during the validation phase, the mock-up was shown to be a fairly good representation of the actual Vision of Man Exhibit, these results suggest that the visuals in the exhibit itself -- artifacts, models, displays, etc. -- make little difference as measured by the tests. However, two related points should be considered in interpreting these results.

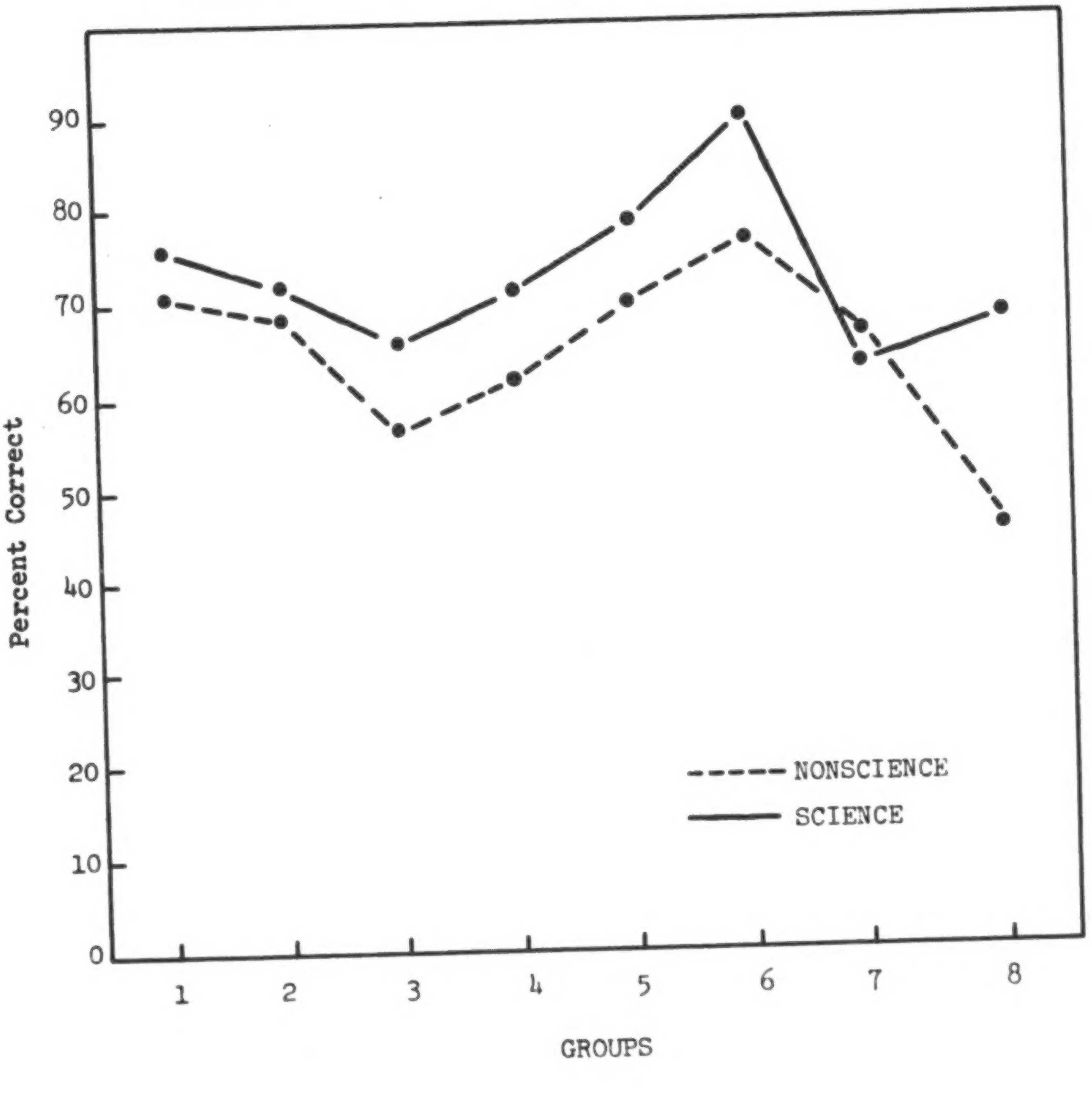


Figure 28. Total knowledge scores (percent correct) of science and nonscience subjects by group in mock-up variation study.

1. The use of total knowledge scores as the dependent variable. The analysis of variance was based on total knowledge scores. The tests used to obtain these scores were for the most part based on the text of the exhibit, not the visuals. The exhibit-only questionnaire (see Appendix J) tended to be based more on visuals than on text, but these items contributed only 7 out of 43 items (approximately 1/6) that made up the total knowledge scores for this phase of work. Even three of these items, #2, #12, and #18 were based entirely on text. However, it can be said that as far as knowledge gained is concerned there appears to be no difference between viewing the exhibit with visuals and viewing the exhibit without visuals under paid, study group conditions.
2. Knowledge gained as an exhibit objective. As has been previously discussed, knowledge gained is only one relevant objective of a science and technological exhibit. As such, it is important to be aware of the fact that the viewer will probably gain most of his knowledge from the text of the exhibit, not from the visuals. However, other objectives -- having to do with interest and attitude -- were not investigated in this phase of the study. The effects of visuals on these more subtle areas may be very important. Additional research would be required to determine the role of visuals in contributing to the formation of, or changes in, attitude and interest levels.

Level of text presented. The analysis of variance failed to reveal a significant difference between the scores of those who received the full text and those who received the simplified skeleton text. A further discussion of this aspect of the exhibit is included in the section of the report dealing with Readability.

Manner of text presentation. The difference between the groups who heard the text and the groups who read the text was found to be significant at the .001 level of confidence. It can therefore be concluded that, in terms of knowledge to be gained, printed signs and labels are better than providing the same content by means of earphones or other aural devices. Again, however, one must ask the question, "How important is knowledge gained in relationship to the effect of sound presentation on other exhibit objectives?" Secondly, a difference of 13 percentage points was measured. The confidence level of the F ratio indicates that 13 percent is a very realistic estimate of the difference that visual versus sound presentation makes. But is this enough of a difference to worry about? The answer lies in the answer to the first question about the importance of knowledge gained. In some exhibits, it may be an important enough factor to warrant choosing reading text presentations over sound; others may choose to include sound presentations in the full realization that some knowledge may be lost, but that other, perhaps subtle, advantages may be gained.

Other operating factors. One can never hope to pinpoint all the factors that contribute to a subject's (or a group of subjects') score on any test. It is for this reason that it is so important to base research projects on large enough numbers of subjects to minimize the importance of unthought-of, unknown, and/or idiosyncratic variations. However, two factors have been shown to be operating on groups of testing subjects often enough to warrant their being considered in this study: IQ and background knowledge. Three analyses yielded the following results:

a. There was a significant difference in IQ between science-nonscience-oriented subjects.
b. There was a significant difference in total knowledge scores between science and nonscience subjects.
c. There was a positive correlation between total knowledge scores and IQ.

The fact that the IQ scores of students who study science are significantly higher than those of ones who don't should not come as much of a surprise and the fact that science subjects had significantly better total knowledge scores than nonscience subjects is also not surprising. In fact, that is why each group was designed to be half science and half nonscience, so that this factor would not raise or lower the score of any particular group independent of the treatment of the group. A question that remains unanswered is whether the science subjects did better because they learned more from the exhibit, or because they had more incoming knowledge. Gain scores would pinpoint the difference, but were not available for this phase of the study.[1]

The methodology. The mock-up variation study was successful as a demonstration of methodology. The use of an experimental design that allows sophisticated statistical measurement of exhibit design variables on less expensive exhibit mock-ups is a step in the direction of a more scientific approach to the entire field of exhibit design. The most important outcome of the study was not the results themselves, but the fact that the design "worked." Because of this, it could be considered a paradigm for further research on exhibit variables, but with the following cautions:

a. More subjects should be included in each group. Ten subjects are frequently too few to find significant differences except in extreme cases. A greater variety of subjects (e.g., lower IQ) should also be included.

---

[1] Control scores obtained in other studies were not applicable here because of the high IQ's of all the subjects who participated in the mock-up variation phase. In order to obtain control scores for this phase, it would have been necessary to run another 80 subjects with comparable IQ's on the test alone.

b. Control scores should be obtained for comparable populations. Gain scores are more meaningful than raw posttest scores, because without the former, it is not clear how much of the test score was due to variations in the experimental variables and how much was due to prior knowledge. (However, this does not invalidate differences observed, only the absolute level of learning.)

Results of The Video Tape Analysis

The collected data consisted of ratings made by three independent judges on the frequency at which design elements were observed by casual viewers in two distinct display areas. Since judgmental ratings are often inconsistent, it was first necessary to verify the reliability of the three judges for each of the displays. The data could not be submitted to further analyses if there was no reliability among judges.

The raw data was summarized from Esterline-Angus tapes. Each numbered position on the tape represented a design element in a display. The small horizontal lines on the graph indicated that an element had been viewed by a casual viewer. For each numbered position, these lines were summed across all subjects for each judge. An individual design element total could show that Judge A thought that X number of subjects viewed position #1 of Fish Sounds X number of times. All individual sums were calculated.

The Spearman Rho rank order correlation technique was used to determine reliability among the judgmental ratings for each display. The Fish Sounds display consisted of fourteen individual design elements. Desalination was made up of eleven elements. The total element sums were ranked from the largest number of observations to the smallest. For Fish Sounds, the rankings ranged from one to fourteen; for Desalination, one to eleven. Rank order correlations were performed between every pairing of the three judges. Table 27 shows the results of these correlations.

All of the correlations between judges were high and were significant. Five of the six were significant at the .01 level. The correlation between judges A and C on Desalination was significant at the .05 level. Although not quite as significant as the other correlations, the ratings of these two judges can still be interpreted as being quite similar. In summary, this analysis indicates that there is a high degree of reliability among judges in their ratings of the number of observations by casual viewers for different design elements.

TABLE 27

Rank Order Correlations and Significance Levels for Judgmental Ratings for Two Exhibit Displays

FISH SOUNDS	SPEARMAN RHO	N	LEVEL OF SIGNIFICANCE
Judge A and Judge B	+.689	14	.01
Judge A and Judge C	+.756	14	.01
Judge B and Judge C	+.831	14	.01
DESALINATION			
Judge A and Judge B	+.855	11	.01
Judge A and Judge C	+.661	11	.05
Judge B and Judge C	+.786	11	.01

Attaining a high degree of reliability among raters was an important factor in establishing the usefulness of the video taping technique as a means for measuring the attracting power of various design elements in an exhibit display. Once the criterion of the reliability was met, there were many possible hypotheses that could have been investigated. The basic approach of this study, however, was to investigate techniques for measuring exhibit effectiveness rather than to evaluate the Vision of Man Exhibit per se.

The wide range in the number of observations assigned to each design element indicates that the levels of "attracting power" were quite different among the individual elements. Some of the elements were looked at by all subjects as many as 37 times while others were not viewed at all. Often subjects would view one element several times while ignoring other parts of the display. Figure 29 shows the Esterline-Angus tape of one subject who was viewing the Fish Sounds display. Notice the many times this subject returned his attention to design element #1, the oscilloscope. He seemed interested in elements #2, #3, and #4, while elements #7 and #12 were examined only once. The remaining eight elements in the display were not even glanced at by this viewer. This viewing pattern was typical of many of the subjects observed and is direct evidence of the great variability in the "attracting power" of the various design elements.

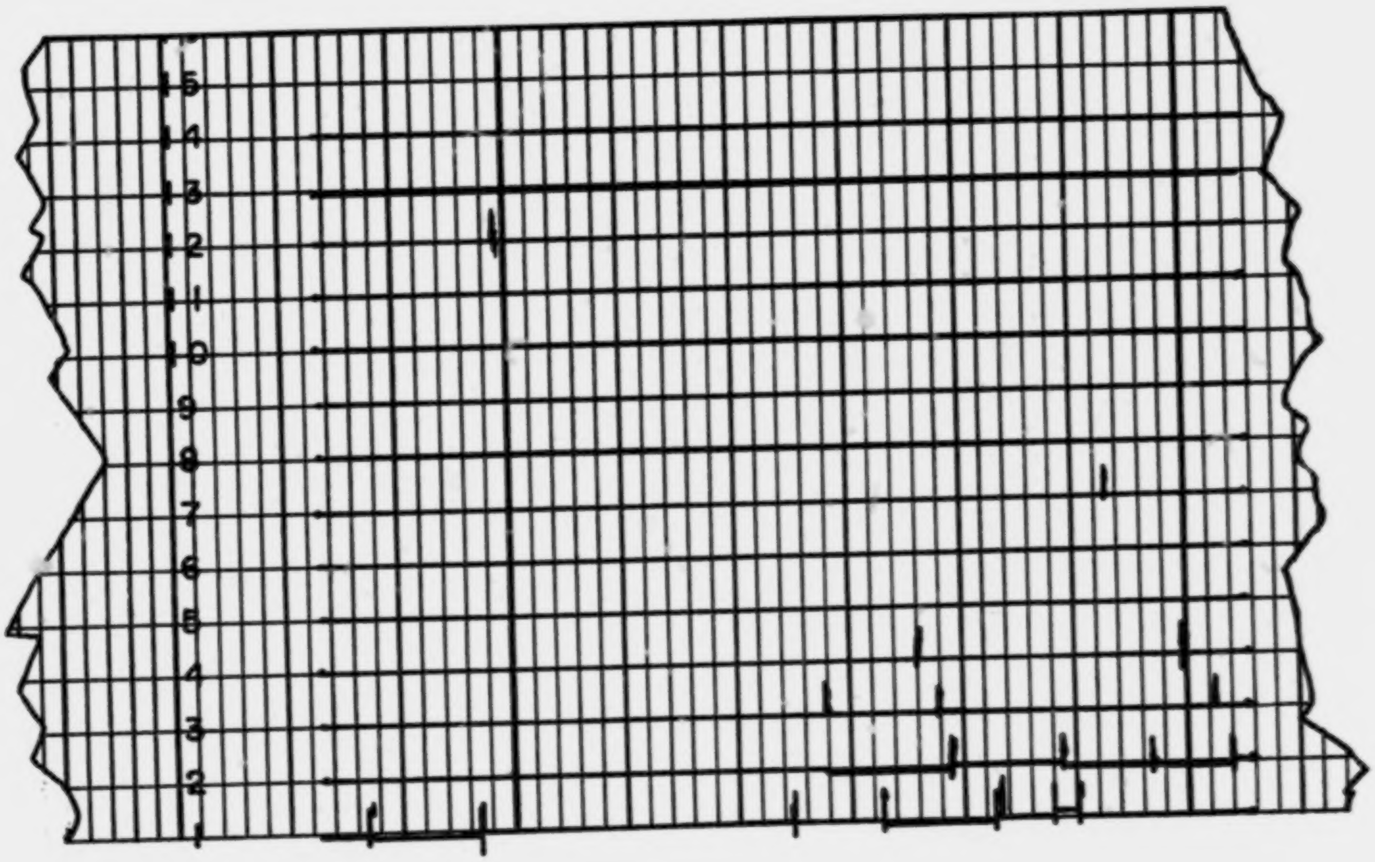
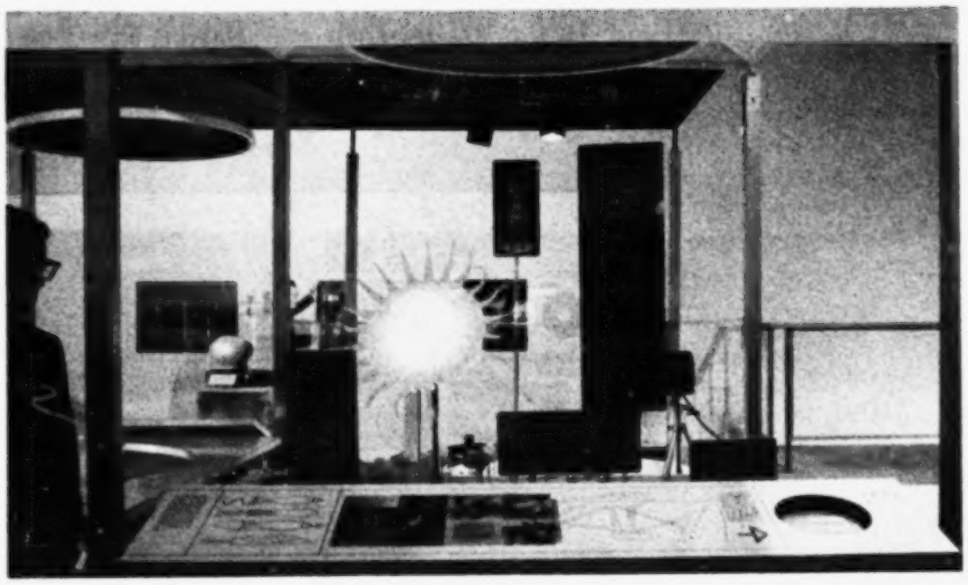


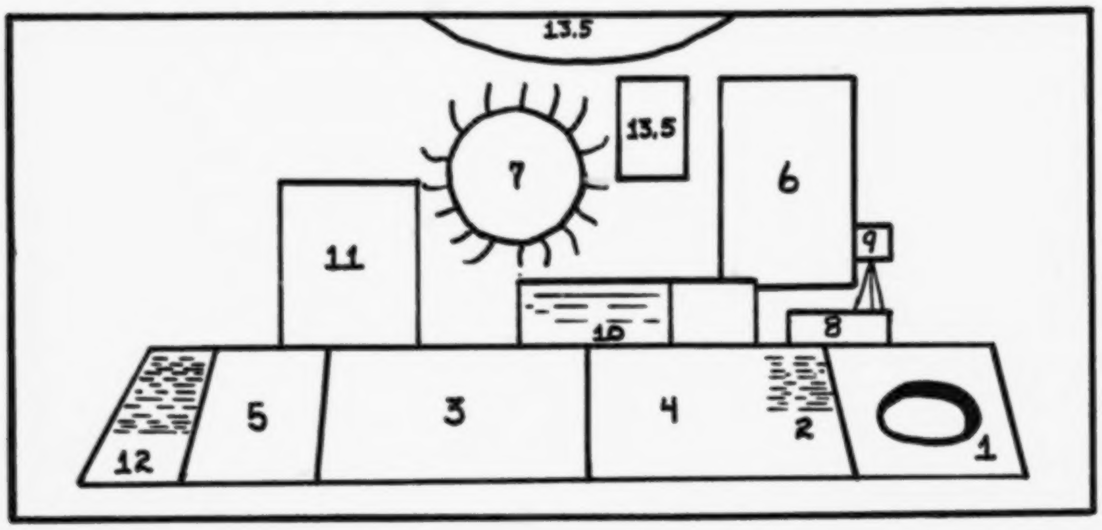
Figure 29. Sample Esterline-Angus Tape, video tape analysis.

A ranked "attracting" scale was developed for each display. The assigned ranks for the three judgmental ratings, obtained from the Spearman Rho correlations, were summed for each design element. These totals were then assigned new ranks ranging from the smallest number to the largest. (It must be remembered that the original rank of one was assigned to the design element with the highest number of observations.) The new rankings established the order of the design elements from most attracting to least attracting for each display.

Figure 30 shows the actual display and, in outline form, the design elements in the "Fish Sounds" Exhibit along with the "attracting" scale ranking assigned to each element. Listed below the figure are short discriptions of each design element.

The design elements ranked numbers one and seven are the two elements in the display that have motion and would thus be expected to rank high on attraction. Element #1, the oscilliscope, was a combination of three dynamic elements. A continuous audio tape of fish sounds was played from the area of the oscillioscope. (The sound, by the way, could be heard for some distance from the display and was noted by the staff as being a very powerful "attractor.") A film alternately showed the fish sounds on a moving oscilloscope tape and a technicolor film of the fish making the sounds. It is interesting to note that the constant design element ranked #2 in the attracting scale was the paragraph describing the film. This sequence indicates that viewers found the film interesting and





1	The film of the oscilloscope patterns and fish along with the fish sounds themselves (audio)
2	Paragraph describing echo-locate principle and the oscilloscope
3	Pictures of porpoises, description of their "vision" and explanation of echo-location and sonar
4	Diagram of ship using echo-locate principle
5	Diagram of fish showing how man learns from nature
6	Tall picture of ocean bottom
7	Model of luminescent jellyfish
8	Small sign which states that scientists are investigating the mysterious sea
9	Television camera
10	Larger sign describing sea animal behavior - salmon migration, etc.
11	Sign describing luminescent jellyfish
12	Paragraph which describes how man hopes to learn more from nature
13.5	Large circular plaque above showing migration patterns
13.5	Small picture of a fish

Figure 30. Video tape analysis, "Fish Sounds" exhibit.

wanted to learn more about it. In ranked position #7 was the other motion element, the model of a luminescent jellyfish. A bright light, placed behind the model, flashed on and off. Although this model could be seen throughout the front section of the entire exhibit, it held only a middle ranking within the Fish Sounds display. Furthermore, the small sign describing the jellyfish was ranked at #11. A separate analysis would perhaps show that the jellyfish model was a factor in originally attracting viewers to the display area but that it was not of sufficient interest to necessitate detailed viewing.

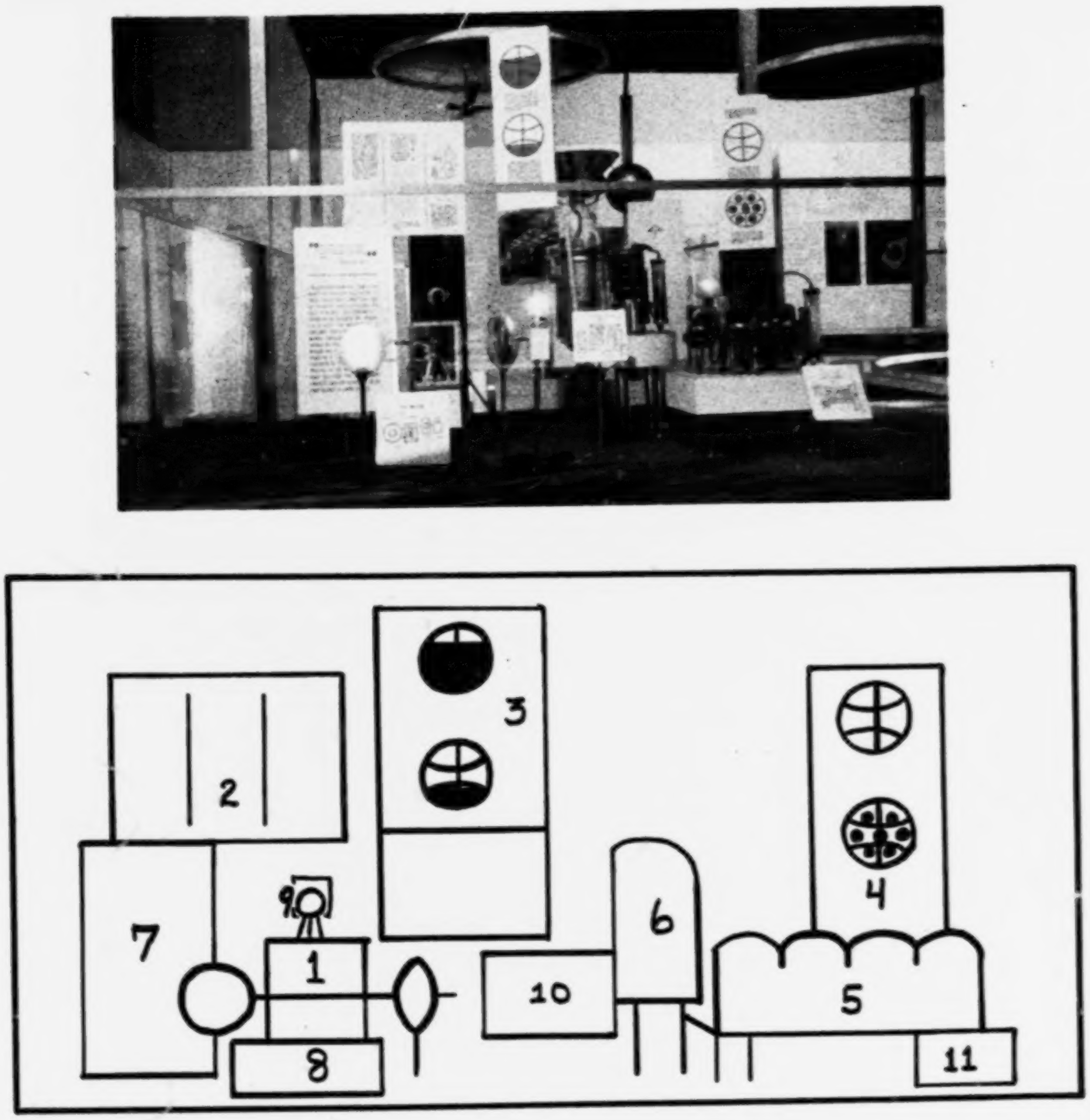
To back up this analysis, a comparison was made between two multiple-choice items dealing with the fish sounds and the jellyfish (Table 28). It seems clear that for all groups, the area concerning the jellyfish produced a lower level of learning than did the area concerning the fish sounds, as indicated by these two items.

TABLE 28

Amount Learned on Two Items Based on Two
Different Kinds of Displays

	FISH SOUNDS	
	MC-2 Fish Sounds %Correct	MC-21 Jellyfish %Correct
Casual Post (N=40)	65	13
Min (N=90)	82	44
Max (N=87)	90	57

The display and the attracting scale rankings for "Desalination" are shown in Figure 31. The dynamic design elements in this display were models of three processes used to desalinate water. The model of an atomic power plant was ranked as the most attracting element in the display. It consisted of moving "machinery" and two types of lights, continuous and flashing. The other two models, multiple flash and reverse osmosis desalting, were ranked in positions #5 and #6, respectively. Diagrams labeling the different parts of the equipment were placed near each model. The diagrams held very low positions in the ranking continuum. Their low rankings may have



Ranked Position	
1	Model of atomic power plant
2	Large sign dealing with 11 state intertie system for power transmission, fusion, and combination nuclear power and desalting plants for electricity and fresh water
3	Tall sign dealing with 7/10's of earth's surface covered with water, most of which is too salty to use and the techniques being developed to assure fresh water
4	Tall sign dealing with energy needs increasing, nuclear energy and fission the answer, and the U.S. has more than 12 nuclear power plants supplying electricity in operation
5	Model of multiple flash desalination
6	Model of reverse osmosis desalination
7	Large sign with President Johnson's quote, introduction to desalination, and cost of energy for desalination
8	Small sign with diagram of atomic power plant
9	The television camera
10	Small sign with diagram of reverse osmosis desalination
11	Small sign with diagram of multiple flash desalination

Figure 31. Video tape analysis, "Desalination" exhibit.

been due to the fact that they were difficult to interpret. Viewers could obtain more information about desalination by reading the display signs than by looking at these diagrams of the models.

An item analysis of an open-end concept and an open-knowledge item covering this area showed that approximately 60 percent of the casual and 85 percent of the study groups knew that desalination was a promising method of combating the world's fresh water shortage (open-end concept), but only 25 percent of the casual and 50 percent of the study groups could name one of the three methods described by the exhibit (open-end knowledge). These results tend to support the rankings made by the video tape analysis. The more detailed and difficult the content, the lower the ranking. This, in turn, is reflected by a failure to learn the material as indicated by the test scores.

Discussion of video tape analysis. The video tape technique offers many alternative methods of analysis. The ranked attracting scale was a rather simple analysis. The obtained results could indicate only generally which design elements contributed to the attracting power of the display. The use of ranks eliminated any discrete differences between elements. Thus, one can not conclude from rankings alone that the oscilloscope in Fish Sounds was twice as effective in attracting viewers as the model of the jellyfish.

If an investigator were interested in evaluating the hypothesis that dynamic design elements attract and hold more viewers to a display than do constant elements, the video taping technique could be adopted specifically for that purpose. An initial step in the investigation would be to select a display with somewhat equal proportions of dynamic and constant elements. The same judgmental rating system could be used in gathering the basic data. Chi square analyses could then be used to determine the differences in attracting power for the two types of elements. Other classifications of design elements could be studied with more complex interrelationships built into the design. Exhibits could be specifically designed and varied to study such relationships with different audiences. A series of studies based on this approach could make a considerable contribution to an understanding of viewer behavior. Studies of this type may be useful in determining the differences between attracting and holding a viewer. A display may be an "attention-getter" because of its overall gestalt, i.e., it is well integrated with each part contributing to a pleasing whole, but may fail to hold the viewer once he begins to attend to the individual parts.

Determining the average time a viewer spends looking at individual display elements and the actual sequence he follows in viewing them could be another use of the video tape technique. The length of viewing time can be easily measured by examining the distance between observations on the Esterline-Angus tape. Notice the length of time the subject observed design element #12 before returning to #1 in the tape excerpt presented in Figure 29. It is apparent that this design element held the viewer's attention longer than did any of the other elements. Time data could be summed across all subjects for each element. These data could be used to answer the following questions:

1. Is the average time spent looking at an element correlated with its position in the overall sequence of viewing?
2. Is the average time spent equal to the amount of time necessary to read the information?
3. Is there a relationship between characteristics of design and the average amount of time spent viewing?
4. Is there a direct correlation between average amount of viewing time and knowledge gained? Knowledge test scores could also be analyzed against the attraction ratings of the elements.
5. Is there a relationship between viewer age, sex, size of group, etc. and viewing time or sequence?

In summary, the video taping technique offers quite a few interesting possibilities as a practical research tool in the evaluation of the effectiveness of individual design elements in a display. The actual taping of visitors and subsequent data collection is time consuming but manageable. Video taping systems and Esterline-Angus Recorders are "off-the-shelf" (but expensive) items. They are available on a rental basis. It is hoped that further work along the lines suggested above may be accomplished with the video taping technique since it has been demonstrated in one instance to be both feasible and reliable.

Results of Readability Analysis

Analysis of data pertaining to the influence of readability on knowledge gained yielded the following results:

1. Mock-up variations. Groups 1, 3, 5, and 8 either read or heard the full text, which was at least one grade level higher than the skeleton text. Groups 2, 4, 6, and 7 either read or heard the skeleton, or lower level, text. The average total knowledge (i.e., multiple-choice + open-end concept + open-end knowledge + exhibit-only) score for groups 1, 3, 5, and 8 was 66 percent.

   The average total knowledge score for groups 2, 4, 6, and 7 was 71 percent. Although the results are in the right direction, an analysis of variance indicates that there is no significant difference between the total full text and total skeleton text groups. (See Table 29.)

TABLE 29

Total Knowledge Scores of Mock-up Variation Groups by Readability

FULL TEXT				
Group	N	ΣX	$\bar{X}$	%
1	10	317	31.7	74
3	10	263	26.3	61
5	10	318	31.8	74
8	10	243	24.3	56
Total	40	1141		
Average	10	285	28.5	66
SKELETON TEXT				
Group	N	ΣX	$\bar{X}$	%
2	10*	301	30.1	70
4	10	285	28.5	66
6	10	358	35.8	83
7	10	280	28.0	65
Total	40	1224		
Average	10	306	30.6	71

F ratio = 2.255 (N.S.)

*One subject omitted some items. He was assigned average scores for those items.

2. Chicago -- High School -- MAX and CONTROL. Items on the four knowledge tests were separated according to the grade level of their referent subareas. For example, multiple-choice item 12 was drawn from the subarea Project Mohole, which had a readability grade level of 8.5 and was thus rated 8.5.

   The items were then separated into low, medium, and high groups on the basis of the readability of their

> referent subareas. All items at grade level 9.9 or below were categorized as low. All items at grade level 10.0 through 10.9 were categorized as medium. All items at grade level 11.0 or above were categorized as high. Gain scores were then computed for each item by subtracting the percent correct for the CONTROL group (high school only) from the percent correct for the MAX group (high school only). The average gain score for low items was 18 percent. The average gain score for the high items was 20 percent. A t test indicated that there was no significant difference between the gain scores on low grade level items as opposed to high grade level items.

Discussion and recommendations. The results of the readability study showed no significant difference in knowledge gained between low and high reading grade levels. Several factors may be contributing to this lack of differentiation.

Subject sophistication. The readability studies were based on high school students only. Except for the Chicago CONTROL group, these students knew that their "job" was to learn as much as they could from the exhibit, and that they would be extensively tested afterwards. It is likely that these students would therefore give easily-read paragraphs only a quick once over. They would devote the majority of their time to material they found more difficult -- memorizing unusual terms, etc. Test experience has taught them that this is the type of material on which they will be questioned.

Typesize. As mentioned previously, 14 different typesizes were used in the Vision of Man Exhibit. There was an indication of an indirect correlation between typesize and grade level. Thus, the phenomenon discussed above may also have caused the subjects to spend more time examining the small print containing the more difficult material.

Legibility. Perhaps the most important factor that can be classified under "legibility" is the placement of the text. The grade level of the Project Mohole subarea is only 8.5, but the text material was placed on the floor, making it difficult to read. The text on Extension has a reading level of 8.7. Not only was the first paragraph located high up on the wall, but the design of the display prevented closer examination. The lack of adequate lighting in this area also made the text difficult to read. In contrast to these two examples, the Theories subarea, which has a grade level of 13.8 (high), was contained in a sign hung conveniently in front of the viewer's eyes. This is not to say that all low level material was inconveniently placed and all high level material was conveniently placed. Placement, lighting, and other factors definitely would contribute to reading ease and thus introduce uncontrolled variability into an analysis confined to readability alone.

The format and spacing of the text probably has some bearing on legibility. The following is quoted from the Radiation Insects and Foods subarea. The format is exactly as it appears in the exhibit.

THE WORLD OF
TOMORROW WILL
FIND NEW AND
IMPROVED CROPS
GROWING ON OUR
LANDS--AND NEW
FOODSTUFFS IN
OUR MARKET BAS-
KETS. BUT WE
WILL STILL HAVE
TO COPE WITH IN-
SECT PESTS &
FOOD SPOILAGE.
FEDERAL SCIEN-
TISTS LOOK FOR
BETTER INSECT
CONTROLS AND
FOOD PRESERVATION
METHODS.

Even though the grade level of this paragraph is 9.7 (low), the choppiness of the lines makes it difficult to read. Other unusual formats found in the exhibit include two-color printing (first line red, second line blue, etc.) and

placing each line
in a separate box
like this.

Nature of test questions. The readability grade levels of the test items themselves were never established. It is obvious that the two questions printed below, both taken from the Multiple-Choice Questionnaire, are written at different grade levels.

23. America's pioneer in liquid-fuel rocketry was:
    a. Marshall Nirenberg.
    b. R. J. Van de Graaff.
    c. Robert H. Goddard.
    d. Richard T. Whitcomb.

27. The subatomic particle which was first found in cosmic rays and is associated with the force that binds the atom's nucleus together is the:
    a. pion.
    b. positron.
    c. ion.
    d. baryon.

Item 23 refers to the Rocket subarea which had an assigned grade level of 11.9 (high). Item 27 refers to the Cosmic Rays subarea which had an assigned grade level of 8.9 (low).

Selection of test items. During the test validation phase, a standard difficulty analysis was done. Items rated as too difficult were eliminated from or replaced in the questionnaire. Although such a technique is common practice in test development, in this case it may have been detrimental to the readability study. If grade level is a factor in knowledge gained, then items that truly pertain to hard-to-read subareas, such as Amplification, would be the most difficult to answer and would be eliminated or replaced during the test development phase.

The use of the Flesch readability formulae. The Flesch Readability Formula is the most commonly used method of determining the grade level of textual material. It is designed to measure the complexity of sentence structure. Unfortunately, other than the fact that difficult words tend to have more syllables, the formula does not consider the vocabulary of the text. In any science/technology display, high level vocabulary words are bound to appear. For example, the following two sentences are quoted from the Radiation-Insects and Food subarea, with a Flesch grade level of 9.7 (low). The sentences themselves are obviously not low.

> *Some insects can be eradicated by radiation sterilization of laboratory-grown young in the pupal stage. ... They were then treated with 8,000 Roentgens of gamma rays to sterilize them.*

The following sentence is quoted from the Cosmic Rays subarea, with a grade level of 8.9 (very low).

> *The subatomic particle called the pion, for example, was first found in cosmic rays, and is associated with the force that binds particles together within the atom's nucleus.*

The following is quoted from the Chemical and Ion Engines subarea with a grade level of 8.5 (very low).

> *An ion engine is being developed in which cesium atoms are heated, ionized, and speeded up electrically.*

Summation by subarea. The grade levels were established by subarea for the entire exhibit. Even casual inspection reveals a great deal of variance within a subarea. (See sentences quoted above.) For example, the text of the DNA subarea presents a variety of information about DNA, using a total of ten paragraphs. Typesize ranges from 1/4 inch to 1-1/2 inches. The average grade level of the DNA subarea is 10.9 (medium). The results of individual Flesch counts for each paragraph in the DNA subarea are presented in Table 30. Although the average readability of this subarea is 10.9, it ranges from 7.4 to 14.2.

TABLE 30

Flesch Grade Level Computation for The DNA Subarea by Paragraph

Paragraph	Grade Level
1. ("This is a representation...")	11.0
2. ("Scientists are unravelling...")	11.4
3. ("Why a man?")	9.0
4. ("Dr. Marshall Nirenberg...")	12.9
5. ("If man can complete...")	10.5
6. ("How did DNA get started?")	10.4
7. ("Some scientists theorize...")	10.2
8. ("Perhaps nucleic acid...")	11.4
9. ("To submit these theories...")	14.2
10. ("It may take generations...")	7.4

An analysis of the questionnaire items drawn from low and high level paragraphs is shown in Table 31. Gain scores were derived by subtracting Chicago high school CONTROL group scores from high school MAX group scores.

These results cannot be considered significant since they were based on a total of two low items and four high items, although a tendency in the expected direction is indicated. It can be assumed that a good bit of the exhibit is similarly arranged and that items now classified by area as high may in fact be medium or low, and vice versa.

TABLE 31

Gain Scores for Low and High Items Drawn from DNA Subarea by Grade Level of Source Paragraph

TEST	ITEM	PARAGRAPH	GRADE LEVEL	CATEGORY	GAIN
OEC	A7	3	9.0	low	30%
EXO	A7	none	very low	low	67%
EXO	A5	2	11.4	high	1%
EXO	A8	9	14.2	high	13%
MC	13	1 & 2	11.0 11.4	high	42%*
MC	45	2	11.4	high	21%

Sum low items = 97
Number low items = 2
Average gain low items = 48%

Sum high items = 77
Number high items = 4
Average gain high items = 20%

*A graphic representation of the four pairs of chemical units accompanied this text. If further analysis of exhibit content yielded results similar to the other three items in this category, it would be logical to assume that the source of the information necessary to answer this question was the illustration, not the text.

With these considerations in mind, one can see that the data collected in this study neither prove nor disprove the hypothesis of negative correlation between reading level and knowledge gained. In order to examine this hypothesis, it is recommended that several steps be taken in future exhibit research studies.

1. Continuation of Flesch (or similar) analysis by paragraph. A rough indication of the influence of readability on knowledge gained can quickly be obtained by continuing the paragraph analysis described above. However, such an analysis would not take into account the other five factors discussed in this section. The control of these five factors must be accomplished by a much more thorough study.

2. Experimental study to pinpoint the effects of readability on knowledge gained. The basic design of this study should be to hold all independent variables except grade level constant. This includes typesize, lighting, attraction, legibility, etc. The dependent variable would be knowledge gained, as measured by paper-and-pencil tests.

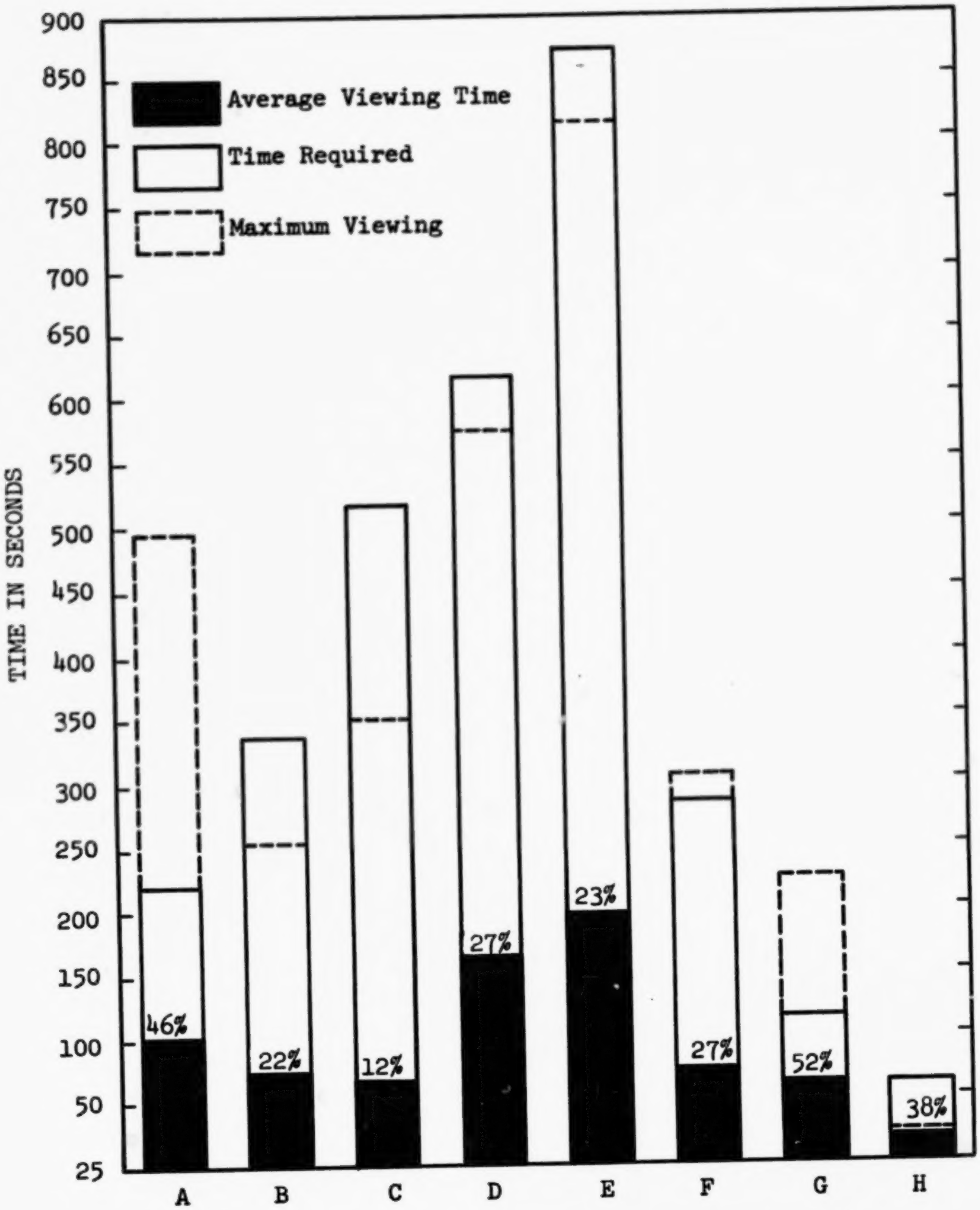
Time Data Results

There are several ways of looking at the casual viewer time data results. Time data can be a general indicator of 1) whether viewers spent enough time to view all of the presented material, 2) the "holding power" of each exhibit area, and 3) how "attracting" subareas are to different age groups.

Figure 32 shows the Chicago casual viewer average viewing time for each exhibit area against the time required to read all of the related materials (at 250 words/minute), including time to view films, listen to a sound track, etc. Note that these data are presented in seconds. The percentages listed show the percent of the total required viewing time represented by the average casual viewer times. These overall percentages suggest that "Careers" and the "Introductory Area" of the exhibit held the casual viewer's attention the longest, i.e., proportionately more time was spent in these areas than other areas. The dotted lines show the maximum time spent by any one casual viewer out of the 60 in the sample. In only three out of the eight areas did even the most interested casual viewer spend enough time to cover the material (assuming he had average reading skills).

Figure 33 shows the subarea breakdown of the casual viewer time data within each exhibit subject area. These times are presented in seconds. (The percentages represent the number of people stopping. This is discussed in the next section.) From these figures, one can see the relative holding power of each subarea.

The total time data collected at each subarea was summed for the entire exhibit and divided by the total number of casual viewers observed. In Chicago, 1989 casual viewers were observed at all exhibit subareas. The sum of their times equalled 40,591 seconds. The average time spent in an exhibit subarea calculated over all subjects was thus 20 seconds. Time data collected in Los Angeles showed that casual viewers spent an average of 37 seconds in an exhibit subarea. One might account for the higher average in Los Angeles on the basis of the relative lack of competing exhibits in the Los Angeles museum (2), as compared with the Chicago museum (many). The time data collected in Chicago were recorded separately for the three age groups, high school, college, and adult.



A - Introductory Area
B - Man, the Living Being
C - Man and His Earth
D - Man and the Basics
E - Man and the Universe
F - Man and the Group
G - Careers and Science and Technology
H - Introductory and Exit Panels

Figure 32. Summary of Chicago casual viewer time data for each exhibit area.

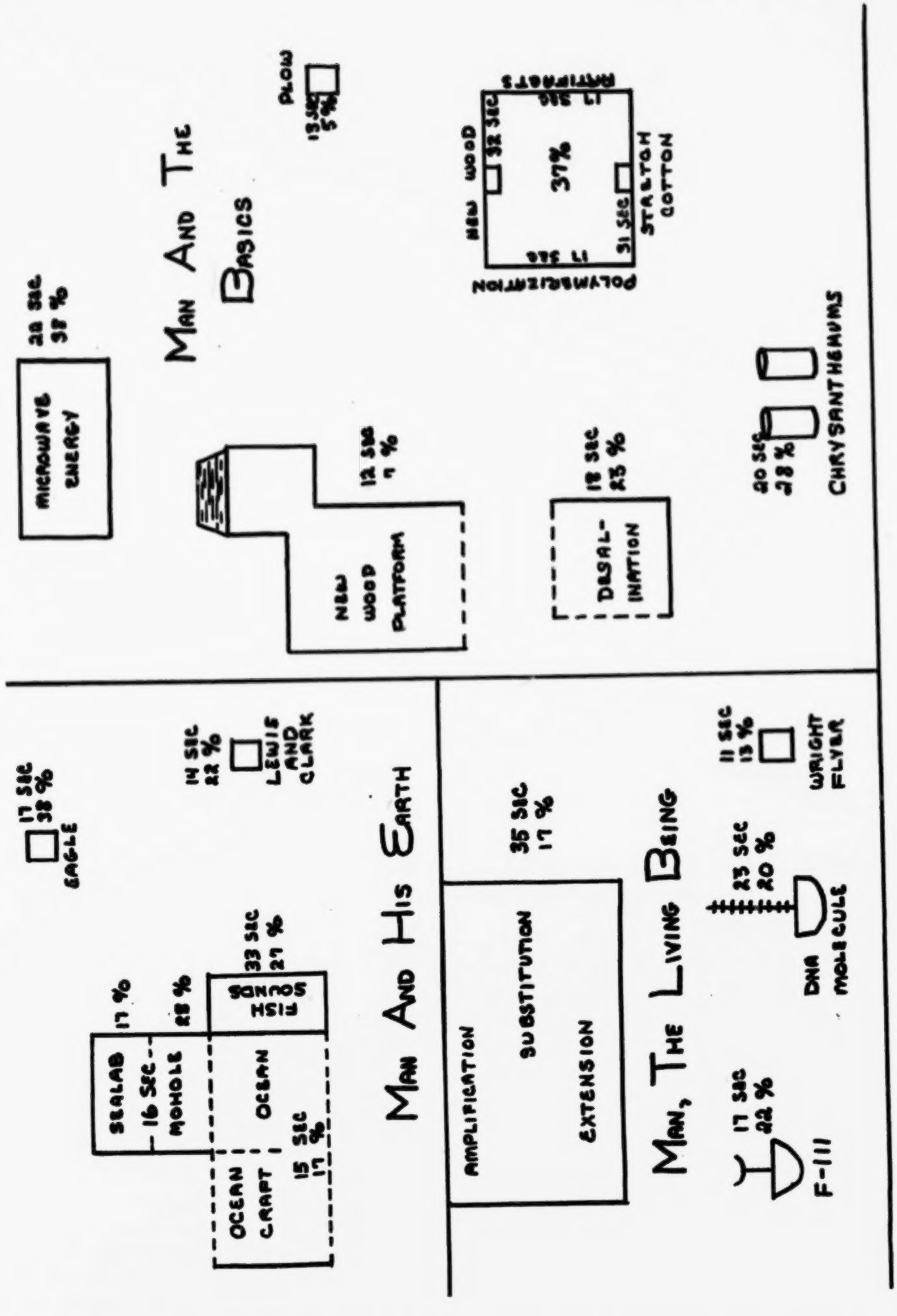

MAN AND THE BASICS
MICROWAVE ENERGY
28 SEC
38 %
NEW WOOD PLATFORM
12 SEC
7 %
PLOW
13 SEC
5 %
NEW WOOD
32 SEC
37%
ARTIFACTS
17 SEC
POLYMERIZATION
17 SEC
31 SEC
STRALTON COTTON
DESAL-INATION
18 SEC
25 %
CHRYSANTHEMUMS
20 SEC
28 %
MAN AND HIS EARTH
EAGLE
17 SEC
38 %
LEWIS AND CLARK
14 SEC
22 %
SEALAB
17 %
16 SEC
MOHOLE
28 %
OCEAN
FISH SOUNDS
33 SEC
27 %
OCEAN CRAFT
15 SEC
17 %
MAN, THE LIVING BEING
AMPLIFICATION
SUBSTITUTION
EXTENSION
35 SEC
17 %
DNA MOLECULE
23 SEC
20 %
WRIGHT FLYER
11 SEC
13 %
F-111
17 SEC
22 %

Figure 33. Exhibit attracting power (in percent) and holding power (in seconds) for each subarea.

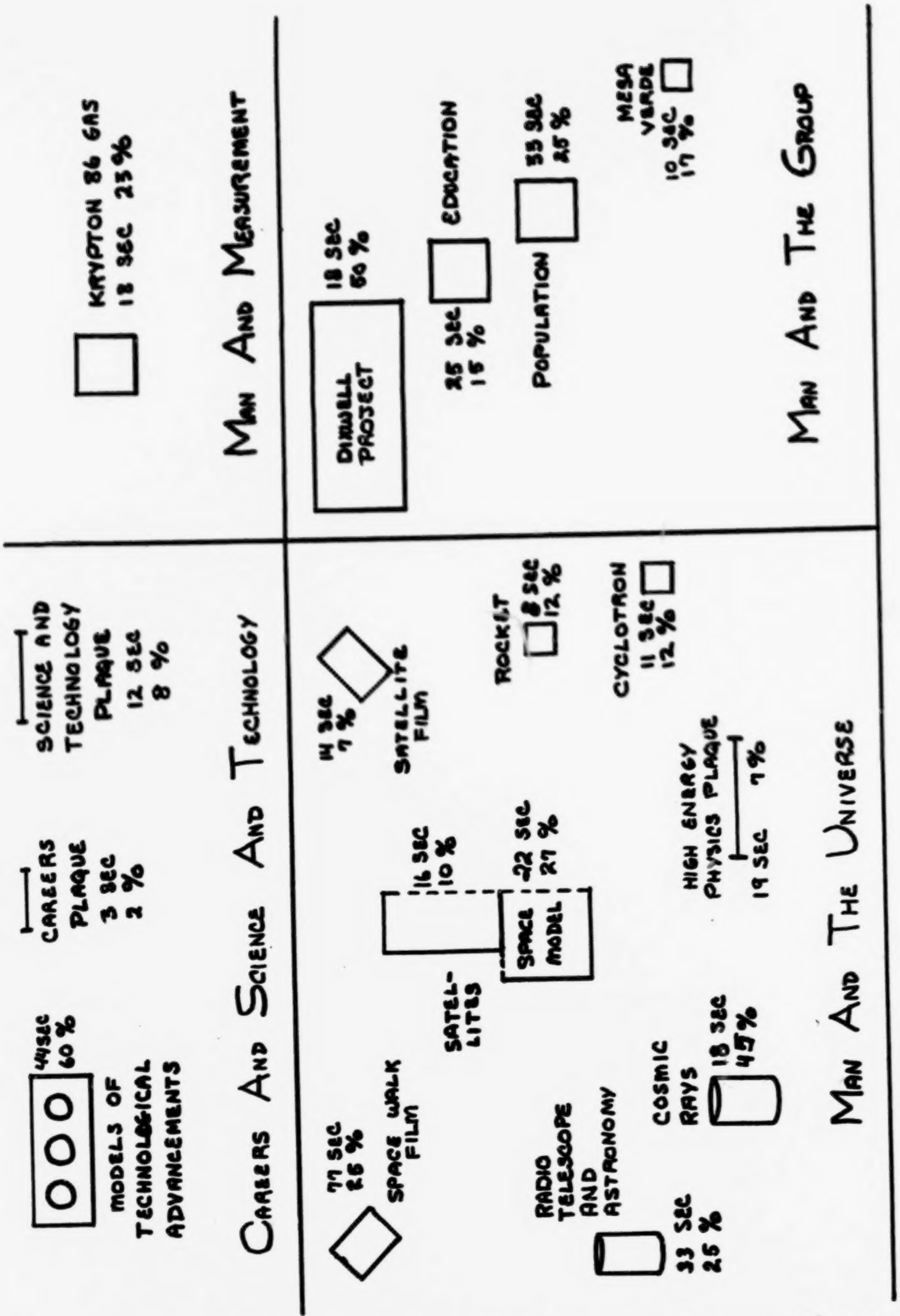

MODELS OF TECHNOLOGICAL ADVANCEMENTS
41 SEC 60 %
CAREERS PLAQUE
3 SEC 2 %
SCIENCE AND TECHNOLOGY PLAQUE
12 SEC 8 %
CAREERS AND SCIENCE AND TECHNOLOGY
KRYPTON 86 GAS
18 SEC 23 %
MAN AND MEASUREMENT
SPACE WALK FILM
77 SEC 25 %
SATELLITES
16 SEC 10 %
SPACE MODEL
22 SEC 27 %
SATELLITE FILM
14 SEC 7 %
ROCKET
8 SEC 12 %
RADIO TELESCOPE AND ASTRONOMY
33 SEC 25 %
COSMIC RAYS
18 SEC 45 %
HIGH ENERGY PHYSICS PLAQUE
19 SEC 7 %
CYCLOTRON
11 SEC 12 %
MAN AND THE UNIVERSE
DIXWELL PROJECT
18 SEC 60 %
EDUCATION
25 SEC 15 %
POPULATION
35 SEC 25 %
MESA VERDE
10 SEC 17 %
MAN AND THE GROUP

Figure 33 (Cont'd.). Exhibit attracting power (in percent) and holding power (in seconds) for each subarea.

Interviewers were instructed to stay in a subarea until 60 casual viewers had been timed. However, for some subareas it was not possible to obtain 60 viewers within the time limits of this testing phase. It might be hypothesized that those subareas with substantially fewer than 60 observations are weak in "attracting" power. Listed below are those subareas attracting less than 60 viewers in the testing phase:

1. Back panel of the Peep Show showing artifacts of new wood and stretch cotton -- 24 people.
2. Wall plaque describing high energy physics -- 5 people.
3. Wall plaque describing uses of radio astronomy -- 18 people.
4. "Score" Satellite Film -- 43 people.
5. Sign describing "new wood" platform -- 32 people.
6. Career opportunities in science -- 2 people.
7. Wall plaque describing science and technology -- 26 people.
8. Exit panel with Federal Government credits for the exhibit -- 7 people.
9. President Johnson's quotation regarding the power of science over nature -- 6 people.

The last four subareas listed above are perhaps the most important topics covered in the exhibit when one considers them in light of the overall exhibit objectives, since they describe careers in science, the interaction of science and technology, and the Federal Government's support of basic and technological research. These objectives were aimed primarily at the high school age group. The two people who looked at Careers were adults. Only four high schoolers viewed the science and technology panel; two, the power of science over nature; and none looked at the Federal Government credits.

Although viewing time can be interpreted only in a general sense, the casual viewer data examined here is discouraging when one considers the time, effort and money involved in designing and building the exhibit. However, further studies would have to be done to determine the relative effectiveness of the Vision of Man Exhibit as compared with other exhibits in this regard. It may well be the case that such data would show that the Vision of Man was superior in both attracting and holding viewers' attention. Without additional studies however, one can only look at the data in terms of the potential of the Vision of Man Exhibit itself.

Number of subjects stopping. These data were collected by interviewers who followed 60 different Chicago casual viewers around the exhibit area (without their knowledge) and noted the subareas in which each viewer stopped. Figure 33 shows the percentage of 60 viewers who stopped at each subarea along with the average time data explained in the previous section. These percentages can be interpreted as general indicators of the attracting power of each subarea. Sixty percent of these casual viewers stopped at the models of recent technological advances (laser beam, pacemaker, etc.). Since this is the highest subarea percentage, it indicates that this subarea was the most attracting part of the exhibit. The second most attracting part of the exhibit, by this measure, was the Dixwell Project showing urban renewal advances in the Chicago area (50 percent).

The smaller percentages indicating low attracting power appear in many of the subareas which were cited as weak in the time data results. For example, the high energy physics plaque attracted only seven percent of the casual viewers; the Score Satellite film, seven percent; wall plaque describing science and technology, eight percent; career opportunities, two percent; and the exit panel with exhibit credits, two percent. [The career opportunities area represents an interesting case of a combination of high and low time data. The overall area data (Figure 32) shows that viewers tended to spend a high percentage of time in that part of the exhibit. But Figure 33 shows that the three models of technological advances accounted for this, while the wall plaque on careers -- a very important part of one of the primary objectives of the exhibit -- was extremely poor in both attracting viewers (two percent) and holding them (three seconds).]

Seventeen of the 60 viewers that were followed by the interviewers started viewing the exhibit at the exit rather than at the introductory area. This confusion developed because the exhibit exit was located near an unrelated film called "Man on the Moon" which was shown continuously. Many people left the film area by way of the Vision of Man exit.

Forty-one exhibit subareas were observed in this analysis. One casual viewer stopped at 24 of these; one stopped at 20; four at 18. The average number of stops for the remaining viewers was 5.7. Fifteen viewers made only one or two stops within the exhibit before leaving the area. These figures contributed to the low percentages shown in most of the exhibit subareas.

In looking at the raw time data results, one is struck by the wide range between the lowest times and the highest times. It seems clear that the casual viewer is extremely fickle. His interest is difficult to arouse and, once aroused, is easily lost. Most viewers passed by the displays without giving them more than a glance. Many others who stopped long enough to be timed, stayed only a few seconds. On the other hand, quite a few viewers who passed this critical level of interest arousal remained for several minutes, but by-and-large failed to remain for the full time required for total coverage of the material. A very few people remained long enough to cover it all. (The video tape materials support this analysis, and further provide an interesting source of "clinical" observation of audience behaviors.) It seems clear that the casual viewer must be attracted by some features of the exhibit that has the capacity to either hold him on its own merits or lead him to something that will. Where interest is initially low, this can present a real challenge to the designer. However, unless this is done, the display will be ignored by a high percentage of viewers.

Determination of Attractiveness Scores

The purpose of investigating this area was to demonstrate the extent to which selected exhibit design variables could be related to one obvious criterion of exhibit effectiveness, "stopping power." Multiple regression techniques were used to develop an attractiveness rating for each of the subareas of the exhibit. Multiple regression is a technique for assigning weights to each of a set of predictors in a fashion that maximizes the correlation between the sum of the products of the predictor scores and their corresponding weights and a criterion variable.

Two design variables were selected as potential predictors of stopping power, the presence or absence of static models, and the presence or absence of dynamic models (models that use motion to illustrate a principle). Not included in the latter category were models that moved but did not illustrate a principle. The F-111 airplane, for example, moved in a circle. Since this movement had nothing to do with the principle of the variable wing, it was considered a static model. On the other hand, the model of the DNA molecule lit up in sequence to show its various constituent parts, and thus was considered a dynamic model. The three measures obtained for each subarea were thus:

1. the number of static models
2. the number of dynamic models
3. the number of viewers stopping at each subarea

Number of viewers stopping was based on an analysis of 60 randomly selected casual visitors who were followed and timed as they moved through the exhibit. "Attraction power" was indicated by the percentage of the 60 visitors who stopped at each subarea. Length of time at each area (holding power) was not included in this analysis.

The three measures were correlated, and the resulting matrix of intercorrelations subjected to the multiple regression analysis. The multiple correlation between the weighted sum of the two predictors and the criterion was +.52, indicating that these two predictors seem to account for, or "explain," approximately 27 percent (.52 x .52) of the variance in viewer stopping behavior. The number of dynamic models in a subarea, however, was a much better predictor of viewer stopping behavior than the number of static models.

The two design variables used in this analysis do not, of course, exhaust the set of variables that could be considered as potential predictors of various criteria of attractiveness as measured by number of people stopping. Lighting, live demonstrations, sound, films, etc., would all be potential candidates for a complete analysis. However, models are generally considered important elements in exhibit design and their expense makes their contribution of more than academic interest.

A few features of the multiple regression technique should be kept in mind.

1. The technique yields the "best" predictions when each of the predictors are relatively uncorrelated with each other, but are correlated with the criterion variable.
2. The weights resulting from the multiple regression analysis should not be regarded as indicators of the importance of each of the predictors since they reflect intercorrelations among the predictors themselves as well as relationships of the predictors to the criterion.
3. Cross validation procedures should be employed whenever possible. After a set of weights have been determined, these weights should be applied to new data as a check on the correctness of the weights.

The actual regression equation for predicting attractiveness is: $A_i = [9.21 + (1.11)S_i + (13.05)D_i]1.666$

where

A_i = predicted percentage of people stopping in subarea i in a fixed time interval

S_i = number of static models in the i^{th} subarea

D_i = number of dynamic models in the i^{th} subarea

9.21 is a constant, 1.11 and 13.05 are B weights and 1.666 is a correction that converts the base of 60 viewers (used in this analysis) to a base of 100 viewers so that percentages of viewers can be predicted instead of absolute numbers of viewers.

Table 32 shows the results of applying the formula (without the correction) compared with the actual number of casual viewers stopping, based on the total number observed (60). Thus, the formula (uncorrected) predicts that 11.4 viewers out of 60 will stop at the Lewis and Clark display case; 23 viewers out of 60 actually did. Similarly, a predicted 11.4 out of 60 would stop at the Plow display; 13 did. Using the correction factor of 1.666 would convert the above figures into percentages.

A replication of this analysis would very likely produce different weights, due to the use of different kinds and numbers

TABLE 32

Number of Casual Viewers Stopping as Predicted by Formula and Actual Number Stopping for Each Subarea (N=60)

Description	Attraction Ratings to Tenths	#Casual Subjects Stopping
Area-Introduction		
Eagle	11.4	23
Lewis & Clark	11.4	13
Plow	10.3	03
Mesa Verde	11.4	10
Wright Flyer	10.3	08
Rocket	10.3	07
Cyclotron	10.3	07
Krypton 86	22.2	14
Area-Man, the Living Being		
DNA	35.2	12
Amplification, Extension, and Substitution	25.5	10
F-111	10.3	13
Area-Man & His Earth		
Fish Sounds, etc.	22.2	16
Ocean (Subs & Fish Migration)	14.7	10
Mohole, Sealab, & Mining	16.9	31
Area-Man & the Basics		
Molecular Δ & New Wood	26.6	22
Agriculture	11.4	16
Desalination & Energy Needs	25.5	37

Description	Attraction Ratings to Tenths	#Casual Subjects Stopping
Area-Man & the Universe		
High Energy Physics	09.2	04
Cosmic Rays	22.2	27
Atomosphere	10.3	15
Space Models	12.5	16
Space Wrench	----	--
Scientific Satellites	10.3	06
Space Walk (Film)	09.2	15
Satellite Film (Score)	09.2	04
Area-Man & the Group		
General	----	01
Education	09.2	08
Population-Chicago	09.2	15
Urban Development-Chicago	22.2	30
Area-Careers & Science & Technology		
Careers	----	01
Science & Technology	23.3	41

of models, different audiences, etc. However, continued use of such a formula would hopefully lead to more powerful instruments of prediction.

It must be emphasized that this analysis refers only to "stopping power" and not to "teaching power." In fact, the correlation between the two was very low, indicating that ability to attract must be followed by effective content to produce total exhibit effectiveness.

Casual Viewer Comments

Comments were obtained from 60 (or 46 percent) of the Los Angeles casual viewers and from 137 (or 43 percent) of the Chicago casual viewers. After completing the test, these viewers were asked the question, "What would you tell your family and friends about this exhibit?" The transcribed versions of these comments were scored as being positive, basically positive with some criticism, negative, or neutral. The results are shown in Table 33.

TABLE 33

Percent of Casual Viewer Comments Categorized as Positive, Basically Positive, Negative, and Neutral

	Los Angeles	Chicago	Total
Positive	70%	61%	64%
Basically positive (with some criticism)	8%	17%	14%
Negative	10%	9%	10%
Neutral	12%	12%	12%

The comments were scanned for most frequently mentioned subjects. The main categories found were:

1. "It was interesting." Many casual viewers mentioned interest in the exhibit, in one form or another. Sometimes this was voiced simply as, "It was very interesting," or, "I found the such-and-such display was interesting." Sometimes viewers mentioned that they thought the exhibit would be interesting to someone else -- "My son is very interested in science and I thought mainly of him." "It would be real interesting for people in college."

2. "I didn't have much time to spend." Numerous people apologized to the interviewer for not having given "his" exhibit the time and attention it so obviously deserved. Many people said they would tell their friends to plan to spend much more time, although the actual length varied from the viewer who thought that "It's worth about half an hour of your time..." to the devotee who would tell his friends and family to "plan on taking the whole day at the Institute instead of just a few hours like I had to." One frank individual commented, "If I were going to tell my friends or family, I'd tell them to be sure and look at it carefully because you take a test after you get out of it." Another offered the suggestion, "I think everyone should be encouraged to go through much slower. I think probably a sign to slow everyone down would be very helpful."

3. "It was very: educational, informative, enlightening." Many people commented on the educational value of the exhibit, often qualifying it with the fact that they did not take full advantage of the opportunity to learn. "It's very educational, but I think I got tired looking at it because it's so big." When you go through an exhibit, you should take advantage of what is there. I did not." "Since I know most of the knowledge in the displays themselves, I went through it fairly rapidly and I had missed part of the exhibit when I took the test." "I'm taking a vacation from school and I'm trying not to learn too much."

4. Comments about the level of the exhibit. These comments fell into two categories:

   a. "It was too: detailed, general," and
   b. "It would have been fine for: college students, high school students, scientists, laymen."

   Also included in this category were those who said, "I didn't understand it." This category contained the widest range of comments, some of which are shown here in contrast:

   "...Personally, Because of a good physical science education, ...I thought that the exhibits on the physical sciences were not informative for me, but very clear for someone who has not had such an education."

   "...But I'd say the majority of the exhibit would deal with somebody that has an interest in a science background and in physics, or something of that sort."

"It's very worthwhile, highly educational, very informative, and presented in a way that even junior high kids can acquire a lot of information from it.	"It was all right for college students, but I didn't like it too much."
"The exhibits are simple enough for a child to understand..." "Bring the youngsters of the family..."	"For the younger child, some of the information is not presented in a way that they could perhaps understand to the best advantage."
"It is especially interesting to the youngsters.."	"I don't think children, you know, should be taken on the tour because they wouldn't understand what the hell is going on."
"I just think there could have been a little more detail given about the subjects."	"They don't just go into it briefly, they go into it fully."
"...it was so general that it didn't give me anything..." "There are a number of points that should have been elaborated on."	"The project should be more simplified, or should have another exhibit for students -- where it's good to give a background or just a generalization."

5. "The whole exhibit was [not] well organized." Several people mentioned the organization of the exhibit. The comments were split among those who felt it was bad and those who thought it was good. Again, the contrast of these comments was very interesting.

"I think the exhibit is extremely well organized."	"...some of the exhibits you can't see."
"I also feel that the pictures on urban renewal and the population explosion were very closely tied together and make it very visual for most people."	"I didn't get any basic feeling of what you're trying to get across. I mean, there's a man here, a heart over there, a whale over there; the connection is almost completely nothing."

"I just want to make the comment what a difference in museums today and when I was a child. They are so interesting and yet they're not cluttered so that you can look at each thing very carefully."

"If you turn around and there's three this way and three this way, it's a little hard to have any continuous feeling for this thing."

Table 34 shows the percentage of people who made comments in each category.

TABLE 34

Percentage of Casual Viewers Who Mentioned Five Most Frequent Categories

	Los Angeles	Chicago	Total
Interesting	45%*	40%	42%
Level	38%	28%	31%
Not enough time	18%	24%	22%
Educational	28%	15%	19%
Organization	13%	11%	12%

*Totals equal more than 100 percent because many viewers commented on more than one subject.

Comments pertaining to organization and to level can be further broken down into positive and negative. Table 35 shows these breakdowns.

TABLE 35

Percentage of Casual Viewer Comments Classified as Level-Positive, Level-Negative Organization-Positive, Organization-Negative

	Los Angeles	Chicago	Total
Level Comments			
Positive	12%	7%	9%
Negative	27%	21%	23%
Organization Comments			
Positive	7%	4%	5%
Negative	7%	7%	7%

The comments were also scanned to see what displays were most frequently mentioned. Negative and positive comments were added together, assuming if someone could make any comments on a display, he or she must have at least looked at that display. The results of this analysis are shown in Table 36.

TABLE 36

Frequency of Mention of Various Exhibits by Casual Viewer

EXHIBIT OR DISPLAY	PERCENT MENTIONED*
Space Exhibits	57
Man and the Group	17
Atomic Energy	10
New Wood	10
Desalination	10
Chrysanthemums	7
Technology	7
Cosmic Rays	7
Each of the following was mentioned once or twice:	
Theories	
Oceanography	
Mohole	
Extension	
Krypton 86	

*Percentages are based on number of exhibits mentioned (42).

Discussion. The summary tables of the casual viewer comments show that 64 percent of those interviewed made positive comments about the exhibit. Another 14 percent made basically positive comments, with some criticism. Only 10 percent made completely negative comments. The comment made most frequently was, "It was interesting." Certainly such data, taken alone, would be very encouraging to the exhibit designer and manufacturer. However, these facts are not consistent with much of the more objective effectiveness data contained in this report. If the casual viewer found the exhibit so interesting, why didn't he spend more time looking at it?

If he found it so educational, why did he score so low on the knowledge questionnaires? In an effort to pinpoint this discrepancy, one must look to "courtesy bias" as a possible solution. The posttest casual viewer had already taken the questionnaires. They told him that the researchers were interested in:

a. his background
b. how much he had learned from the exhibit
c. how interested he was in the exhibit

Being courteous, he would like to tell the researchers that he found their exhibit interesting and educational. Many people did just this, and did not attempt to develop this point. Examples:

> "Well, I would tell them it's very interesting and that they could learn a lot if they had never come here before and didn't know about certain things; that they could really learn a lot by the exhibits that they have here. That's about it."
>
> "Yes, I think that the exhibit is very interesting and instructive, and I think that it's well worthwhile to view."
>
> "I thought it was kind of enlightening. It was interesting."
>
> "It's an interesting exhibit."

However, most people found it necessary to attempt to elaborate on the value of the exhibit, and this is where one can begin to spot courtesy bias. Some elaborations were quite successful, and making a convincing case for that person's interest in the exhibit. Examples:

> "I liked the whole exhibit in general because we covered a great different variety of subjects which you are going to stimulate someone's interest, no matter who they are. I think there is something here that would interest just about everybody. It is set up so that the children will be interested in it as well as the adults. I have a great feeling for science in general, so there were two or three things that interested me a great deal. I liked your thing on the mutation of plants. I think you ought to make that a little bit bigger than what it was. That stayed in mind because that's what I took in high school and college. I'm interested in plants and animals. I thought it was a good diversified exhibit."
>
> "For myself, I'm not really that scientifically oriented, but the part that interested me the most was the idea of population increase. I'm interested more in social attitudes, coping with the masses. Insofar as what I would

> tell my family, I would very highly recommend it to my father to bring the kids in because I think children should be exposed to what's going on in all phases, knowledge generally."

> "I thought it was an interesting exhibit from the fact that it showed me things that I had read about before and had never actually seen, and here you got to see them and sometimes you actually got to work them yourself. I though it was really good."

Unfortunately, many comments fell apart in the elaboration phase. The following examples are split into "courtesy" comment and unsuccessful elaboration.

"I would say it was very interesting.	Mainly, I came to push buttons, I think there should be more gadgets. And I guess things that really interest people are things that say 'push this button'."
"I would tell them it was interesting,	but I couldn't exactly say why. There was so much to see that I couldn't assimilate it all in one 30- or 40-minute period."
"I would say it was all right.	I don't really know what it was about."
"I thought the exhibit was rather interesting,	I don't know. The only reason we came down was for something to do, and, I don't know -- I'm not really too interested in this."

Perhaps the most frequently occuring pattern was the casual viewer who attempted to excuse himself for not having been very interested, personally, in the exhibit or for not having learned very much. The courtesy bias played a big role in these comments. In one type of comment, the courtesy bias took the form of "I didn't learn too much [or, I personally wasn't interested], but I'm sure someone else would [be]." Examples:

"I'd say the majority of the exhibit would deal with somebody that has an interest in a science background and in physics, or something of that sort. [There] seemed to be a small smattering of everything, and I think it would be interesting; ....but as far as I'm concerned, there were just a very few topics that were really interesting that I paid attention to."

"My son is very interested in science and I thought of him mainly, and then my husband would be interested too. ....I know so little about it that it goes in one ear and out the other."

"I don't have the interest or the background in science to answer questions about pros or cons. ...

....

:..Someone that did have the interest would enjoy it."

"It was all right for a college student, ...but I didn't like it too much."

In another type of comment, the courtesy bias took the form of an excuse. The most common excuse, of course, was, "I didn't have enough time." Examples:

"First of all, I think it was important that we went through the exhibit pretty hurriedly, ...but from what I saw, it was very gay and it did arouse interest. I would recommend it to anyone who came here."

"Although I went through the exhibit very quickly, ...I was impressed with the general intention of the exhibit to display the intentions of modern science. I would recommend anyone to go to the exhibit. It was very educational."

"I think it is a real good exhibit, and I think everyone should come and see it; ...

....

..but I for one haven't had enough time to really look over it real good, ...but I think it's a pretty good exhibit."

"I would say it's an interesting exhibit, ...

....

..but I don't know about its good points or bad points because we didn't have much time because we only have a few hours here and then we have to leave right after lunch, and we are trying to see as much as we can."

"The exhibit was very interesting and all, ...

....

..and as I went through, I didn't take time to read all the little things, so the questionnaire was kind of hard to answer. ....In general, I thought it was very interesting and quite informative."

Excuses other than time were offered, some of them rather unique.

> "If I were really concerned about it, I would have taken a greater interest, but it was just -- I came to see." [sic]
>
> "I went through kind of fast, but I think if I saw it, I would have liked it a lot better."

Perhaps the most important outcome of the interviews was the few constructive criticisms made. Some of the examples are listed below:

> "I think that as a whole, it's a pretty good exhibit, but you try to switch its topic too many times -- like you see one small area and you look at that and then all of a sudden you switch from treating wood to outer space and things like this; and so I think it would have done better to go deeper into maybe a fewer number of subjects and not to skip around so much."

"It was a fairly nice exhibit, but I think you should have more on the recent moon shots. More exhibits of the actual models. I would like it if you had information handed out at a desk like pictures, etc. Other than that, it was a very nice exhibit."

"It is quite complex. Actually, I think you should have some plan of how people should walk through it -- a certain type of way -- like arrows or something springing from the simple to the more complex. It was really a very interesting exhibit."

"It was a good, clean, basic exhibit, but I think it lacked action. The more moving parts that it has, the more moving exhibits, the more action it will attract to the most people. I think it is set back out of the way of the Museum, and it isn't in a prominent place and no one can get to it. I think if you are going to have an exhibit like this, you should have it where people can get to it and not just by accident."

"Quite good, I think the exhibits on manufacturers of those products are missing. I mean there should be more on the manufacturing side also."

"The only thing I think needs to be improved is the starting and stopping point. I think I have come in on the middle of it or missed some if it; I'm not sure because I didn't know exactly where I was going."

"For one thing, I think there should be automation and visual aid rather than looking at pictures. Other than that, I think it's a pretty good exhibit."

"It was interesting, but if you put it a little simpler, you would get more people to understand it."

"I thought parts of the exhibit were informative, but it didn't tie together -- the whole concept of whatever this new science is. For instance, when you walk in, there should be some kind of announcement or some type of eye-catching device that will enable you to know what you're getting into, because as you go further and further you see all these different sciences the oceanography and space, etc., you wonder how they do tie in."

"I'm really not sure what I looked at. I think there should be an explanation in the beginning, maybe there was and I missed it. I was walking through and really looking for another section. I was more or less lost when I stumbled into it and after I got looking at it,

> I was interested in parts of it (the cosmic rays and the conservation of our resources) but after having been there, I think I should go back in there and take another look, because after reading these questions, I see a lot that I missed that I really wasn't paying attention to when I started in there. Needs an explanation in the beginning."
>
> "I think it's a very fine exhibit. I think everyone should be encouraged to go through much slower. I think probably a sign to slow everybody down would be very helpful."
>
> "I just sort of glanced at the exhibit, but what I didn't like about it was visually I didn't think it was presented right -- too many sounds, too many giant things going along with it. It should have been more of an approach where you would be drawn to it visually and more interested, rather than just a hodgepodge."

In summary, interview data of the sort reported on here tends to be "soft" and "fuzzy." On casual inspection, it may even be downright misleading. When analyzed, however, such data can serve a useful purpose. In fact, these comments, along with the video tapes of casual viewers actually looking at the displays, may be more meaningful to exhibit designers than all of the "hard" data presented in this report. Unless exhibit designers design exhibits to please other exhibit designers (and we know that isn't ture), then the comments and actions of the real "customers" of their services ought to be noted with genuine interest.

DISCUSSION AND CONCLUSIONS

Reading the previous 143 pages of this report may create a reaction similar to that produced when a young school boy was told to write a report on a book about penguins. His report was short and to the point: "This book told me more about penguins than I wanted to know." However, an exploratory study should attempt to "tap" as many areas as possible; in fact, it would not be at all difficult to make a list of measurement variables not considered in this study, or ones given inadequate attention, or variables whose interactions were not explored.

This final section of the report will review the work that was accomplished, emphasizing the problems encountered and highlighting the results that show the greatest promise. Some of this material has been covered in earlier sections of the report in the context of individual studies. No effort will be made to duplicate here the level of detail contained in these previous discussions. This section will also present a general theory of exhibit effectiveness in an effort to provide a conceptual framework for future studies in the area.

Objectives

This topic is listed first because it occupies a position of primacy in exhibit effectiveness studies. Stated very simply, the results obtained in any evaluation study of an exhibit can be no better than the statement of objectives for that exhibit. The difficulties experienced in the present study in obtaining objectives suitable for evaluation purposes is not felt to be unique. If only one step could be taken to improve the "state of the art" in exhibit evaluation, it would be to state exhibit objectives as clearly, concisely, and "behaviorally" as possible.

This does not mean that such objectives must therefore be trivial, although the insistence on "behavioral objectives" in the field of programmed instruction often has this undesirable result (18). "Music appreciation" sometimes becomes "memorize 20 composers' names, birth dates, etc., etc." General and even lofty aims and goals are useful for establishing an overall direction, or universe of discourse. However, such statements must be restated in terms of *specific viewer behavior that should be brought under the control of the exhibit.* Thus, the objective "the viewer should have an increased appreciation of Federal science" must be analyzed further into subobjectives, i.e., the specific items of behavior that would be evidence of such "appreciation." The designer could then select the materials ("hard" and "soft") that would be most likely to lead to such behavior, perhaps even to the extent of requiring the viewer to practice making overt responses consistent with the subobjectives.

[The reader not familiar with the notion of "behavioral objectives" is encouraged to refer to a small book by Mager, "Preparing Objectives for Programmed Instruction" (16) for an introduction to this area.]

The impact of this problem area on methodologies for exhibit evaluation is profound. If the objectives of an exhibit are vague and ambiguous, the method selected for evaluation may be inappropriate. Those responsible for the exhibit can then take refuge from possible negative effects of such evaluations by saying, "But that is not what we were trying to do." The instructional "loop" must be closed before evaluation can be placed on firm ground. (This "loop" was shown on page 3 of the Introduction to this report.) Note that in the present study, little emphasis was placed by the exhibit sponsors and organizers on learning per se, i.e., on the assimilation of factual knowledge on the part of the viewer. The project staff inferred knowledge objectives from the general statements of objectives, and from the fact that the exhibit itself contained a vast amount of factual information. Since this study was designed to evaluate methodologies and not the Vision of Man Exhibit itself, an error in judgment in this regard would not invalidate the study. If, however, it was the purpose of this project to evaluate the Vision of Man Exhibit, then the emphasis placed on knowledge gained may be considered inappropriate, and the results misleading.

The task of preparing objectives for a single concept, didactic exhibit would appear to present the least difficulty. An exhibit study currently in progress is concerned with just such a display, designed to teach young children (between 5 and 15 years of age) the relationship between tooth shape and function in animals.[1] The test materials for such an exhibit can be tied in directly with the teaching objectives (e.g., discriminate between different kinds of teeth). Thus, this exhibit can be treated (and, in fact, is being treated) solely as an instructional device, with all of the procedures currently used in the development and validation of such devices being directly applicable.

Greater difficulty would be experienced in taking this "instruction" approach to an exhibit designed for less tangible and immediate results, e.g., exhibits designed to "create interest," or to "increase appreciation." Scientific and technical exhibits are often a mixture of both and probably occupy a position between pure didactic and "inspirational." At the far end of the continuum would perhaps be art and craft exhibitions, designed solely to "please the eye." However, the original conceptualization stated in the introduction to this report, and carried through the individual studies, still applies. All exhibits should be considered essentially as educational media. This approach makes their evaluation a technical problem (although a difficult one) and not a metaphysical problem.

---

[1] This is a USOE, Bureau of Research supported project entitled "The Development of Validated Museum Exhibits," directed by Elizabeth Nicol. Information was obtained through personal correspondence.

Methods of Evaluation

As noted above, the appropriate method to use for exhibit evaluation depends on the objectives of the exhibit. The other major consideration is the purpose of doing the evaluation. The present study showed many of the possibilities for measurement, including paper-and-pencil measures of knowledge, interest and attitude, and nonpaper-and-pencil measures having to do more directly with viewer behavior. A comprehensive evaluation ought to employ most or all of these types of measures, but practical considerations may preclude this in many situations. When this is the case, selection should be based on the most relevant measures in terms of the key or primary objectives. Thus, the didactic exhibit might best be evaluated solely by paper-and-pencil knowledge tests, and not be concerned with attracting power, interest, and other, less tangible elements. On the other hand, an exhibit designed specifically to "stimulate interest" might best be evaluated solely by an interest measure plus analysis of viewer behavior (video tape or direct observation of attracting and holding power).

It should be noted that the present study attempted to base the development of its various measuring instruments on a comprehensive analysis of the exhibit content. This approach provides considerable diagnostic power for determining the relative contribution of the various individual elements within the exhibit. If the purpose of the evaluation is to improve the effectiveness of a particular exhibit or to increase the store of knoweldge re exhibits in general, this comprehensive approach would be appropriate. If, however, concern is limited to how well the overall objectives of a particular exhibit have been achieved, then measures related only to those objectives would be more appropriate. Naturally, the comprehensive, diagnostic approach is recommended here as providing the greatest long-term benefits, ultimately and hopefully making it possible to predict exhibit performance, rather than relying upon empirical effectiveness measures.

Measurement studies with actual exhibits must also be supplemented by laboratory studies to provide more basic, fundamental knowledge, as was done with the mock-up variation study. This is suggested by the fact that exhibits exist in a bewildering variety, and field studies, even ones as comprehensive as that described in this report, are always subject to the specific conditions and limitations that surround a particular, existing exhibit. Further suggestions for such laboratory studies will be covered in a separate discussion.

A final general comment on evaluation measures and their treatment. One of the most difficult problems in the area of exhibit evaluation centers around experimental design requirements. For example, the simultaneous consideration of a number of different variables and their interactions necessitates attention to the problem of equalizing cell frequencies in any multi-variable classification scheme. In a two-way classification scheme, employing three levels of the first variable and four levels of the second variable, a total of twelve cells result (3 x 4 = 12). As the number of levels of each variable increases, and as the number of variables increases, the total number cells or individual experimental conditions mounts dramatically since the number of individual experimental conditions or cells increases as the product of the number of levels in each variable. Thus the simltaneous consideration of more than three or four independent variables having more than three or four levels per variable becomes impractical. Requirements for subjects become unmanageable and the results uninterpretable. The fact that there are both conceptual and statistical limitations to such analyses, however, does not make the phenomena under investigation any less complex. Exhibits are complex, and do involve the interaction of a large number of variables. Research strategy must be adapted to these realities. The studies reported on here utilized some of these more advanced techniques. Future studies must consider analysis of variance and covariance design, for example, as part of the standard armamentarium of exhibit research.

Mock-up Validation

The mock-up study provided considerable evidence that useful validation work and research work could be accomplished using the mock-up approach. It is clear, however, that the adoption of mock-up prevalidation techniques on the part of exhibit sponsors and designers is not likely to be immediate or enthusiastic. Time and cost factors would probably be the two primary expressed concerns, but it is also safe to say that many of those working in the exhibit field are simply not convinced that prevalidation is needed. Tradition tends to be against treating exhibits as objects of rational investigation, and tradition is not easily ignored. The educational field still has not completely accepted the concept of prevalidation, despite the fact that educators are much closer to the influences of educational and psychological research than are those who work with exhibits. But change can come about, (and, to some extent, is coming about) when the consequences of inaction are made clear. A mock-up prevalidation study of Vision of Man, for example, similar to the mock-up validation study described in this report, would have pointed out many areas where useful changes could have been made. There is no reason to believe that similar results would not be obtained for other exhibits. When the notion becomes more fully accepted that exhibits are capable of systematic investigation, their prevalidation, using relatively inexpensive mock-ups, may be the rule rather than the exception.

Audience Variables

The Achilles heel of exhibit research is the audience. In the final analysis all the efforts of the sponsor, the designer, and the fabricator will be to no avail if the audience does not come to, or stop to see, the exhibit, or, having seen it, is not influenced in a direction consistent with the objectives of the exhibit. But the exhibit viewer is not well understood. On the basis of the present study, he seems to be easily distracted and not easily instructed. Perhaps the most disturbing of all the results reported on in this study are the dismal scores obtained by the casual viewer group on the various knowledge tests. The casual viewer appears to be fickle, jumping about from display to display, looking for something that interests him. Watching him "in action" on the video tape monitor at the Vision of Man Exhibit, he seemed to be challenging the exhibit to "make him stop," or to "give him something to be interested in." Of course, he can also be an intensive observer, devouring every last word, picture, model, etc., in the display. This happens, however, much less frequently.

It will perhaps never be the case that the exhibit designer can "match" his exhibit perfectly to audience characteristics. However, much more attention needs to be given to the viewer, even if only on a gross level. Ultimately, it may be possible to say a good deal more about exactly how various exhibit elements like reading level, size of print, color, length of text, sequence of items, size and type of models, etc., should be expressed to match various audience variables. Fifteen years of research in programmed instruction has not provided all the answers (or even very many of them) to similar questions, and the "audience" for these materials is generally much more homogeneous, and, in addition, is "captive." The solution to the problem of inadequate knowledge that has been applied in the field of programmed instruction has been the use of tryout and revision procedures. This reverts back to the use of mock-ups to prevalidate exhibits with the intended audience. When one isn't sure whether or not the (behavioral) objectives have been achieved, the only way to determine it for certain is to test the audience. At the same time, hopefully, more basic research work will continue so that better predictions can be made and prevalidation may no longer be required. Such predictions, however, are very much in the future.

In the meantime, the audience remains as the sole source of justification for spending any money at all on the preparation of an exhibit. If nothing else could be done in the way of measurement or prevalidation, it would be worth the time of any exhibit designer to simply talk to exhibit viewers. The comments obtained from casual viewers in the present study are a gold mine of ideas and criticisms. Such data, of course, tend to be "soft" and "clinical" and do not lend themselves to statistical analysis and predictions. However, the several rather simple steps taken to organize the comments obtained from casual viewers provided a number of consistent findings that probably accurately reflects attitudes and opinions. A basic premise of the programmed instructional approach to teaching is that the learner is "never wrong." If a student cannot correctly respond to a step in the program, it is taken as an indication of failure on the part of the programmer, not the student. While the "connection" between a casual viewer and an exhibit is not as direct as that between a student and a program, the principle is still valid. The reader is invited to re-examine the comments obtained from the casual viewers of Vision of Man (pages 133 through 143) to see how they might be used as a basis for exhibit modifications.

Retention

Retention is the abandoned child of most educational and media research. Everyone would like to ignore it, but it is still on the door step, waiting to be recognized. Exhibit research is no exception. Of what value is the demonstration of an increase in knowledge or a change in attitude immediately after exposure to an exhibit if these effects are transitory? The answer most often given is that the researcher must start somewhere, i.e., if he hasn't demonstrated learning to have occurred shortly after exposure to the medium then obviously there is nothing to be retained. Thus, it might seem fruitless to carry out retention studies when shorter range studies indicate that little initial knowledge is being gained from an exhibit (as appeared to be the case for the Vision of Man Exhibit in this study). The logic behind this argument may be sound when applied to factual knowledge. The same approach, however, is somewhat more risky when applied to interest and attitude data. It is entirely possible that short term follow-up studies may indicate that attitudes and interests do not appear to have been influenced by an exhibit when, in fact, the long-range effects on these variables might turn out to be quite powerful.[2]

---

[2] It should be noted that the interest measures used in this present study showed considerable change from control group to experimental group, while attitude measures showed very little change. Expressions of interest thus seem to be amenable to change, while expressions of attitude reflect a more stable quality.

The classical studies of Hovland and his associates on "sleeper" effects in attitude formation point out the importance of obtaining attitude and interest measures at different points in time following the exposure to materials (14).

Although longer range follow-ups may be more important in the areas of interest and attitudes, their application to knowledge measures should not be ruled out entirely. The phenomenon of "reminiscence," in which more rather than less of the original material is recalled after a time interval, has been demonstrated frequently in laboratory learning experiments (2). The material that is recalled after a prolonged passage of time, however, tends to be more general than that recalled initially.

The existence of "reminiscence" in the knowledge domain and "sleeper" effects in the interest and attitude domains both point up the need for follow-up studies conducted at different time intervals following the exposure to an exhibit. In addition, these two phenomena have implications for the kinds of measures that should be taken to evaluate an exhibit, and, indeed, may have implications for the design of the exhibit itself. Since the passage of time and the intervening events that take place during this interval may be expected to produce systematic modifications of the original effects of the exhibit, some sort of "synthesis" would be expected to take place in which specific facts, pieces of information, and impressions become "congealed" into a more or less general sort of "cognative blob" or gestalt. Thus, viewers might be able to express general principles or feelings at some time following their exposure to an exhibit without being able to "defend" them with specific facts or incidents. If this is the case, the kinds of questions asked at later intervals should be sure to include those of a general nature. If this phenomenon can be demonstrated, then the choices of specific pieces of information for inclusion in an exhibit itself should be guided by an awareness that the information per se is likely to be forgotten by the audience. What will remain in the viewer's mind is a summary, an overall impression, or a set of conclusions drawn from the specific information presented. Hopefully, this general cognitive picture will be the one which the exhibit designer intended to produce, i.e., consistent with his objectives. A basic research question that suggests itself in this context is the relative effectiveness of exhibits that stress factual information and de-emphasize the general, broad principles as compared with exhibits that de-emphasize the specific and emphasize the general.

In summary, measures of interest, attitude, and knowledge should be collected over a period of time in order to determine the more lasting effects of exhibit viewing and their interaction with exhibit content and design variables.

Laboratory Approaches

The number of possible laboratory approaches for simulating the features of an exhibit is obviously quite large. The mock-up studies described earlier in this report represent one approach to the problem of exhibit evaluation. Another approach that might be used will be briefly described below. It is flexible in its application although it constrains the viewer even more than did the mock-up variation study. It is intended only as one example of what could be usefully done to advance basic knowledge in the exhibit field.

Subjects are seated in front of a rear view projection screen upon which the stimulus materials are presented. The viewer can exercise control over the materials by manipulating a hand held switch which is connected to control and recording devices. Photographic slides of exhibit materials (existing _or_ specially prepared) are placed in a slide projector in back of the viewing screen. Control devices would be used to determine the rate at which slides are presented and thus the duration of the individual slide presentations.

For example, the control apparatus could be set to present the slides at the rate of say, one every two seconds unless the subject is willing to "work" in order to keep a slide on the screen longer by depressing a switch. By connecting this switch to a rheostat on the projector and to a timer in the control apparatus, the projected image will gradually fade away after an initial preset period unless the subject is willing to depress the switch to maintain the image. It would be assumed that the amount of effort that a subject is willing to exert in order to view a particular slide above and beyond some minimum exposure period is correlated with, or is an indicator of, the interest value of that particular slide.

The information that could be obtained in a laboratory set up such as this is limited only by the imagination and ingenuity of the experimenter. Audience variables could be systematically manipulated to see what patterns emerge. Text materials could be studied as well as pictures, models, etc. Sequencing strategies could be investigated.[3] Sound and motion could be studied in a more elaborate configuration. And, finally, automatic data recording devices would simplify that aspect of the work.

---

[3] A USOE supported study of sequencing strategies found that a multiple-concept strategy, in which simple descriptions of several related concepts covered at the beginning of instruction, is more effective than a single-concept sequence, in which one concept at a time is covered in its entirety (22). Studies of this sort should be replicated to determine their transferability and validity in the exhibit medium.

In summary, the potential usefulness of laboratory research in increasing basic knowledge of exhibit variables is quite high and may be the most desirable approach in view of the economics governing the construction of large exhibits and the complexities of doing controlled studies in the field.

The Role of Theory in Exhibit Research

Research obviously cannot take place in a vacuum. It requires a focus, a starting point, perhaps even an underlying theoretical framework. The primary purpose of theory is to guide future research and to integrate past findings. An exhibit theory cannot attempt to embrace all the complexities of exhibits at this time (and for a long time to come). Such a theory must therefore deal with a few variables that are thought to be of overriding importance and attempt to spell out the mechanism by which they interact. The theory need not have a high degree of formal elegance. It should, however, produce testable, nontrivial, hypotheses that can serve to guide research. It should change from time to time as new information becomes available.

A theory of exhibits, or any theory for that matter, must be judged in terms of the amount and quality of competently performed research it stimulates, and not necessarily in terms of its ultimate "explanatory" properties. That virtually any theory is at least partly "wrong" in one way or another is the nature of science. The scientific method is as destructive as it is constructive, for each new theory is built upon the "ruins" of earlier ones and is forced to reckon with the empirical facts and relationships unearthed by its predecessors. A theory is useful when it tells one what to look for, where to look for it, and what to expect to find -- explicitly.

The field of exhibit design and evaluation seems to be in need of the systematic influence and guidance that such theory can exert. The actual form the theory should take is not of particular importance so long as it stimulates meaningful investigations. The theory, once developed, will be forced to change, regardless of its starting point, in order to account for empirical facts and relationships uncovered by the research it has stimulated and guided.

It is quite likely that some persons in the field of exhibit design will misinterpret this emphasis on theory construction and verification as an attempt to reduce the creative aspects of exhibit design to a set of scientific formula devoid of creative or artistic merit. No such results are intended, nor would such a course of action be wise. The emphasis on developing a theory of exhibit effectiveness is simply an attempt to make explicit the decision rules followed by those who design exhibits in order that these rules may be subjected to empirical verification, and thus to modification and improvement.

A Three Factor Theory of Exhibit Effectiveness

This report will conclude with a description of a three factor theory of exhibit effectiveness. It is derived from the results of the present study, and is consistent with the educational orientation taken throughout this work. However, it also reflects a characteristic of exhibits that sets them apart from other educational media, i.e., the voluntary nature of the audience. Exhibits generally do not have "captive" audiences; other educational materials generally do. There are, of course, exceptions on both sides. Didactic, school exhibits may, in fact, have a captive audience since they are often viewed by a class as a specific assignment, and tests are often given on their content. On the other hand, a classroom television lesson or film may be optional, and thus qualify as "voluntary." However, these exceptions do not invalidate the basic point. Exhibits, by and large, are unique among educational media in the noncommitment, or lack of "contract" (implied or otherwise), between the audience and the medium. A theory of exhibit effectiveness that fails to take this essential fact into consideration cannot presume to represent, or account for, the total exhibit medium. The casual viewer time data (viewers stopping and average time per stop) and knowledge data obtained in this study should be recalled by the reader who doubts that this is true.

A comprehensive theory of exhibit effectiveness must thus concern itself with three areas: 1) initally attracting viewers to the exhibit, 2) maintaining their attraction throughout the exhibit, and 3) maximizing the amount of relevant learning or "influence" that is achieved on the part of the viewer. If an exhibit is weak in any of these three areas, the chances of its achieving its stated objectives would appear to be greatly lessened. To restate the three factors in a general way, an exhibit must exercise sufficient control over the viewers behavior so that he is attracted, he stays and he learns.[4] The sponsor and designer must decide in advance upon the specific responses or criterion behaviors that the viewer is to make and then, applying principles of behavior control, select the appropriate means for producing these behaviors in the intended audience.

The first item, i.e., initially attracting the desired audience to the site of the exhibit, is in part a publicity/advertising problem; the second item, i.e., getting the audience to stay at the particular exhibit until it has been viewed in its entirety, is similar to "point of purchase" advertising in that the members of the audience must often select from a number of different but proximal alternatives. At present, solutions to problems relating to "attractiveness" are far from being found. They remain largely in the hands of "experts" who would admit that they "fly by the seat of their pants" all too frequently. Although it is not a particularly difficult problem to determine whether or not a publicity or advertising campaign has succeeded in attracting exhibit audiences having the desired characteristics once it has been opened to the public, it is difficult to decide, in advance what the potential success of various alternative approaches will be. The generation of audience appeals is thus a creative process in which a strictly rational problem solving procedure has so far not been applied. The same sort of statement may also be made with respect to the "point of purchase" aspects of the problem, i.e., after the viewer has reached the site, how does one influence him to view a particular exhibit which must compete with other exhibits sharing the same general location?

---

[4] "Learning" is used here in the context of behavioral change as it has been used throughout the study.

The present study suggests several hypotheses related to this area that could be explored in future projects:

1. Open sound (i.e., not private sound, such as earphones) is a potent stimulus and will generally pre-empt other stimuli. It is long range in its effects and thus could be recommended for general crowd attraction. However, used to excess, it becomes self-defeating (as do most other attracting "devices").
2. Motion is also a potent attractor, but more at the point of application.
3. Age and type of intended audience will influence what works best in attracting and keeping an audience. The younger the audience and/or the less their pre-interest level matches the exhibit content the greater the emphasis on attraction devices and techniques.
4. Interest level can be measured and used to help determine what techniques should be used for different content areas (e.g., the "best" attractors for the lowest interest areas, etc.).
5. Over attention to interest and attention factors will be self-defeating. The audience will tire of them and/or they will interfere with the real purpose and goals of the exhibit.
6. Audience participation has high attraction value but may have low instructional value unless carefully programmed and controlled. Children are extremely easy to "capture" with such methods, but not easy to instruct. Watching children race from

"push button" to "push button" without so much as glancing at the explanatory legend is a sobering sight to anyone concerned with the educational value of the exhibit medium.

7. High strength attraction mixed with low interest, poorly organized or difficult content, will result in a high dropout rate among viewers.
8. Learning per se is interesting to most people. Thus, things that help them learn could be considered as inherently "attracting."
9. Viewers are constantly engaged in a series of trade-off decisions, comparing present "value" with alternative "values." Interest level factors in part determine the direction the decision takes ("stay" or "leave") and the extent or depth of commitment to the exhibit (i.e., "stay to the end" or "stay to see just the next display, and if it's no better, leave," "skim the rest," etc.).

Many other hypotheses could be listed, but these will serve to indicate the relationship between the first part of the general theory and the kinds of hypotheses that can be generated and tested.

The third factor in the theory has to do with the attainment of exhibit objectives once the audience has been "captured." Exhibits may be "programmed" in much the same fashion as other more traditional programmed instructional materials. A set of principles governing the preparation of programmed materials has been built up over the years on the basis of practical experience. Although these principles do not qualify for the label of "theory" in the strictest sense, they do represent what is felt to be a useful conceptual framework for exhibits, particularly those with clearly educational objectives.

Unfortunately, but understandably, the principles of programmed instruction have tended to neglect the very area that has been noted with concern in the previous discussion, i.e., attracting an audience and maintaining high interest levels. The writer of programmed learning materials is usually justified in assuming that the user of his materials has a certain level of interest in, or motivation to learn, the tasks required of him. In fact, the program user could be assumed to have the same essential set of objectives as the program writer. Consequently, one can assume that "getting the correct answers" and "advancing through the program" are sufficient as reinforcers. The typical user of programmed materials is frequently constrained by the situations governing the occasions of their use. He may use them as part of a formal course of instruction in a school or college, or as part of a training program in a military or industrial situation. In all of these environments, the reinforcement contingencies are usually structured so as to assure continuation and completion of the program -- or else! The "voluntary" purchaser and consumer of programs may also be assumed to have sufficient intrinsic interest and motivation to complete the program.

The typical exhibit viewer, however, should not be assumed to share the same learning objectives as the exhibit designer. Most typically, all that can be assumed is that he has a certain amount of time and a small amount of (latent) curiosity. For these reasons, the translation of the principles found to be effective in programmed instruction must be made with great care lest they detract from, rather than contribute to, the development of more effective exhibits by making the exhibit dull, boring, and/or pedantic. With this in mind, the following principles found effective in preparing programmed materials are presented as hypotheses related to the third factor in the theory.

1. Exhibit objectives must be stated explicitly and carried out in the conception of the exhibit. Every sign and label, every model or mock-up, every slide or movie, and every button to push and lever to press, must be related to an exhibit objective or subobjective. (Since one objective of any exhibit would almost inevitably be concerned with attracting and keeping the appropriate audience, interest and attraction devices can be included in the list of items "related to an objective.") The selection of appropriate instructional media is a difficult process; it is far from clear "what media ought to be used to achieve what kinds of objectives," as a recent USOE study has shown (1). But some rational thought can still be given to the relationship, as Gropper has shown in developing television science lessons (10, 8, and 9). Is motion an intrinsic part of the objective? Is color? What is important about a particular relationship or conceptualization? How is it going to be "shown"? In short, the exhibit must appear to the viewer as having a theme and a coherent unity, with all of its parts contributing in a specific way to that theme. When this is not done, viewers will sense the lack of "direction" and will reflect it in relative lack of interest and attention.

2. Exhibit content should be presented in small steps, or increments. The size of the step is determined by the characteristics of the audience and the difficulty of the subject matter. In principle, however, the step should be no larger than can be "taken" by the average viewer. Gilbert has conceptualized this principle in his approach to programming called Mathetics (6). He calls this unit of material the "operant span," which is the maximum amount of information that can be retained by the learner before

reinforcement or knowledge of results is introduced. The concept of step size can be applied at the overall exhibit level and not just at the display or "label" level. A common fault of exhibits is that they try to do too much. Item 1 above would tend to restrict the scope of an exhibit since it would reduce the likelihood of trying to achieve unrealistic or unattainable goals. This item further reinforces this notion by forcing attention on the viewer's "capacity to learn." (The interest shown recently in the notion of single concept films is an outgrowth of this approach to the film medium. "Single concept exhibits" may be a useful conceptualization in the exhibit medium. Even if a series of such exhibits were put together to form a larger display, the individual units would reflect, and benefit from, the "behavioral objectives, small step" approach.)

3. Exhibits that make maximum use of relevant, overt responding on the part of the viewer will be better instructional devices than those that do not. Responding done in a relevant context is better than responding isolated from such a context. (One projected exhibit design is actually called a "response box." Many channels of two-way communication between viewer and exhibit are suggested in this design.)

4. Exhibits that inform the viewer of the correctness of his responses will produce a higher level of responding than those that do not. This item and item 3 suggest "game" devices as part of the display. Such devices not only create the opportunity for overt responding and knowledge of results, but also suggest possibilities for automatic recording of responses for evaluation purposes.

5. Exhibits that sequence the content in a rational way will be better instructional devices than those that do not. Sequencing can be based on several models, e.g., easy to hard, general to specific, temporal (i.e., early to recent), inductive, deductive, geographic, etc. Several of these may be mixed, but the best sequence is the one that the viewer can recognize, i.e., the logic behind the sequencing strategy ought to be clear (or made clear) to the viewer or it isn't a useful strategy. Repetition is a useful instructional principle that can be made a part of the sequencing plan.

6. Crowd control is an important exhibit element that can "make or break" the usefulness of the previously noted items, but particularly item 5. General overcrowding, bunching up in "bottle necks," or lack of provision for viewing by more than one person at a time, are several

common types of problems. Exhibit effectiveness is directly related to the success with which such problems have been handled.

7. Exhibits that have been prevalidated with their intended audience (in mock-up or full configuration) and revised on the basis of the results will be more effective than those that are not. This procedure can be used to help determine the extent to which the previous six items have been adequately handled. It is thus an "escape hatch," preventing errors from becoming a permanent part of the display.

8. As a refinement of item 6, exhibits that attempt to appeal to a wide (heterogeneous) audience ought to make special and separate provisions for these various levels. A "track" scheme is suggested whereby the low interest "skimmer" can proceed rapidly through the exhibit and obtain objectives relevant to his level, the medium interest, average viewer can proceed more slowly and achieve his appropriate objectives, and the intensely interested "digger" can satisfy his craving for knowledge and reach objectives consistent with his level of interest. Color coding and size and placement of type are several methods for cuing such a track system so that the viewer can follow it. Readability and content analyses similar to those conducted in this study would help to determine the appropriate material for each track. Failure to do this results in an exhibit that is too hard for some, too easy for others, and just right for those who represent the middle of the audience distribution. This approach to general public exhibits should increase learning and decrease exhibit "drop outs."

9. The first six steps can be combined into a general "control" hypothesis, i.e., the exhibit that exercises the most relevant stimulus and response control over the viewer will potentially achieve the greatest impact on the viewer. However, the first two factors in the theory (i.e., "attracting" and "holding" the viewer) must be accounted for in this formulation, lest the viewer consider the exhibit aversive and seek to escape from the control being exerted. This leads to a final hypothesis that combines elements from several of those already stated: *The best control methods and devices are those that are both effective and the least obvious, i.e., that produce the appropriate responses on the part of the viewer in the context of what appear to be voluntary actions.*

Naturally, the relative weight applied to the various factors noted above would depend upon the nature of the exhibit and a lengthy list of environmental constraints. For example, didactic exhibits can exert more obvious control than inspirational exhibits. Traveling exhibits have limitations that do not apply to permanent exhibits. Space available is a factor. Low budget exhibits cannot use the complex and sophisticated devices that are available to exhibits supported by large budgets. "High control" exhibits must be particularly concerned with crowd flow patterns, solutions to which can be both difficult and expensive. In this connection, high control exhibits tend to reduce exhibit capacity and this presents the exhibit sponsor with something of a trade-off dilemma, i.e., large audience, low instructional value -- small audience, high instructional value. These and other "practical" considerations could be optimized only in the context of the particular "mix" of factors operating in a given situation. Before such decisions can be made, however, a great deal more must be learned about the matrix of variables that constitute the raw material of exhibits. The present effort, and others, can at best be described as useful probes. Systematic, replicated studies must be undertaken, some done in the field to "shake down" the real issues, and some in the laboratory, where the necessary controls can be exercised. In this way, a gradual improvement in exhibit effectiveness can be realized. It will take not only time, but funds. Those who benefit from the use of exhibits must decide on the priority for this work.

If the present effort can be said to have one major strength, it would be in its demonstration that exhibit complexity is subject to rational investigation and that current educational practice in the areas of both measurement and instructional technology hold out the promise of real advances in exhibit effectiveness.

REFERENCES

1. Briggs, L. J., Campeau, P. L., Gagne, R. M., & May, M. A. *Instructional media: A procedure for the design of multimedia instruction, a critical review of research, and suggestions for future research*. Pittsburgh: American Institutes for Research, 1967.

2. Bugelski, B. R. *The psychology of learning*. New York: Henry Hold & Company, 1956.

3. Calver, H. N., Derryberry, M., & Mensh, I. W. Use of ratings in the evaluation of exhibits. *American Journal of Public Health*, 1943, *33*, 709-714.

4. Carmel, J. H. *Exhibition techniques--traveling and temporary*. New York: Reinhold Publishing Corporation, 1962.

5. Fine, P. A. The role of design in educational exhibits. *Curator*, 1963, *6*, 37-44.

6. Gilbert, T. F. Mathetics: II. The design of teaching exercises. *The Journal of Mathetics*, 1962, *1*, 7-56.

7. Gropper, G. L., & Lumsdaine, A. A. *Studies in televised instruction. Report #7. The use of student response to improve televised instruction: An overview*. Pittsburgh: American Institutes for Research, 1961.

8. Gropper, G. L. Why is a picture worth a thousand words? *AV Communication Review*, 1966, *11*, 75-95.

9. Gropper, G. L. Learning from visuals: Some behavioral considerations. *AV Communication Review*, 1966, *14*, 37-69.

10. Gropper, G. L. Programming visual presentations for procedural learning. *AV Communication Review*, 1968, *16*, 33-55.

11. Grove, R. Museums come alive. *American Education*, 1967.

12. Gulliksen, H. *Theory of mental tests*. New York: Wiley, 1950.

13. Hess, H. H. Attitude and pupil size. *Scientific American*, 1965, 2-10.

14. Hovland, C. I., Lumsdaine, A. A., & Sheffield, F. D. *Experiments on mass communication*. New Jersey: Princeton University Press, 1949.

15. Krugman, H. E. In the eye of the beholder. *Sponsor*, 1964.

16. Mager, R. F. *Preparing objectives for programmed instruction.* San Francisco: Fearon Publishers, 1962.

17. Reed, V. D. *Report and recommendations on research methods used to determine the impact of and reactions to the U. S. official exhibits in international trade fairs*. Washington, D. C.: Office of International Trade Fairs, U. S. Information Agency, 1957.

18. Shettel, H. H. Objectives, "ours" and "theirs." *National Society for Programmed Instruction Journal*, 1964, *III*, 12-14.

19. Shettel, H. H., & Reilly, P. C. *An evaluation of existing criteria for judging the quality of science exhibits*. Pittsburgh: American Institutes for Research, 1965.

20. Shettel, H. H. *An evaluation model for measuring the impact of overseas exhibits*. Pittsburgh: American Institutes for Research, 1966.

21. Shettel, H. H. *Atoms in action demonstration center: Impact studies*. Pittsburgh: American Institutes for Research, 1967.

22. Short, J., & Haughey, B. E. *An experimental study of sequencing strategies*. Pittsburgh: American Institutes for Research, 1966.

23. Taylor, J. B. Science on display: *A study of United States science exhibit Seattle World's Fair, 1962*. Seattle, Washington: Institute for Sociological Research, University of Washington, 1963.

24. Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. *Unobtrusive measures: Nonreactive research in the social sciences*. Chicago: Rand McNally, 1966.

APPENDIXES

APPENDIX A - Story Line for Exhibit on Federal Science and Engineering for Museum of History and Technology Exhibit Objectives
APPENDIX B - Fact Sheet on the Federal Science and Engineering Exhibit - The Vision of Man
APPENDIX C - Physical Layout of Los Angeles Exhibit
APPENDIX D - Posttest Background Questionnaire
APPENDIX E - Post-Interest Index
APPENDIX F - Attitude Questions
APPENDIX G - Open-end Concept Questions
APPENDIX H - Open-end Knowledge Questions
APPENDIX I - Multiple-choice Questions
APPENDIX J - Exhibit-only Questions
APPENDIX K - Physical Layout of the Mock-up Variation
APPENDIX L - Summary of Background Questionnaire Results
APPENDIX M - Individual Cell Frequencies and Means for the Analyses of Variance

APPENDIX A

APPENDIX A

STORY LINE

for

EXHIBIT ON FEDERAL SCIENCE AND ENGINEERING FOR MUSUEM OF HISTORY AND TECHNOLOGY

EXHIBIT OBJECTIVES

The objective of the exhibit is to fire the imagination of young people about the impact of science and technology on the world. We want to awaken them to the exciting current accomplishments in science and engineering and to stimulate them to think seriously of pursuing studies and selecting careers as scientists, engineers, and technicians. We want to encourage many of them to teach in these fields. We want to show that in ... THE WORLD OF FEDERAL SCIENCE AND ENGINEERING ...

ARE -- the thrill of exploring, the delight of discovery, the pride of conquering, the heroism of pioneering, the challenge of building, the satisfaction of service

ARE -- invention, innovation, mystery, suspense, adventure

ARE -- your future, your Nation's future, humanity's future!

In developing this world of Federal science and engineering we want to show the interrelationship of basic and applied science and technological development and its impact upon man ...

We want to show the fantastic future, such as the exploration of space, the stuff of the earth's crust, the exploration of the oceans, the breaking of the genetic code ...

We want to show the practical side, such as frozen foods, standard dress sizes, new industries created, the development of new drugs and their testing to combat diseases ...

In presenting this world of reality and wonderment, we want to emphasize that scientists and engineers are an essential part of our society. We want to show them as flesh-and-blood individuals coming from every stratum of our society and not white-smocked, colorless thinking machines whose scientific and technological pursuits demand absorption and prevent individuals from living well-rounded lives.

BACKGROUND

The Federal Government has been concerned with science and technology since the beginning of the United States of America. As a matter of fact, Article I, Section 8, of the Constitution provides: "The Congress shall have the power ... to promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors and the exclusive Right to their respective Writings and Discoveries." But this rather inconspicuous nod to science in the Constitution does not reveal the extent to which the idea of Federal Government encouragement of science loomed in the minds of many of our Nation's founders.

At the time of the founding of our Nation, the relationship of science and technology and the impact of one upon the other was not really demonstrable. Generally, inventions were the result of the ingenuity of craftsmen working on a trial and error basis rather than the result of careful application of scientific theory. However, the usefulness of science had gained acceptance. Physics and astronomy were proving of value to ships' captains and military men, and natural history had a close alliance with medicine. Furthermore, the body of scientific knowledge had not grown so great nor so specialized that its limits were beyond the reach of educated persons. Knowledge of natural history and natural philosophy was part of their culture.

The Constitution and Science

As educated men, the framers of the Constitution were familiar with science and valued it. (Benjamin Franklin, for example, was one of the great scientific men then living.) Their knowledge and appreciation of science and a belief that the new Nation had a responsibility to diffuse knowledge among its people led them to give consideration to the constitutional position of science in the government they were trying to form.

As a result of Article I, Section 8, of the Constitution, the first patent law, designed to encourage individual ingenuity and secure for inventors some benefits of their creativity, was enacted in 1790; the Patent Office itself goes back to 1802, before there was a clear distinction between the philosopher and the scientist. In the same year, engineering was given emphasis by the Federal Government when President Jefferson created a Corps of Engineers which "shall be stationed at West Point in the State of New York and shall constitute a Military Academy."

First Government Scientific Activity

The first actual scientific activity of the Government dates to 1807, when the Congress authorized a survey of the coast and established the Government's first technical bureau -- existing today as the Coast and Geodetic Survey. Along with the Coast Survey, early governmental involvement in scientific activity took the form of sponsorship of ventures such as the Lewis and Clark expedition, which made significant findings in botany and zoology, and S.F.B. Morse's testing of the telegraph. Other landmark actions included the establishment of an Office of Weights and Measures (forerunner of today's National Bureau of Standards), the legislative requirement that the Patent Office test each invention (calling implicitly for the use of scientific principles), the creation of the Naval Observatory and of the Smithsonian Institution -- all of which came about before the mid-19th century.

As new departments and agencies were established in later years, the enabling legislation began to include specific provision for a scientific function, sometimes reflecting congressional intent that the department or agency become the repository of the most authoritative information related to the missions of the organization. The classic example is the 1862 act creating the Department of Agriculture, which requires the agency "to acquire and diffuse ... useful information on subjects connected with agriculture in the most general and comprehensive sense of that word, and to procure, propagate, and distribute among the people new and valuable seeds and plants."

Permanent Bureaus and Scientific Functions

The Department of Agriculture's creation marked the beginning of the era of permanent bureaus with scientific functions. The next 48 years saw the establishment of a growing number of such organizations, among them the Signal Corps, the Naval Hydrographic Office, the Fish Commission, the U. S. Geological Survey, and the Weather Bureau.

With the 20th century and the emergence of large-scale industry as the dominant force in the Nation, a new type of Government scientific organization came into being, in response to the developing close relationship between industry and science. Noteworthy among these organizations were the National Bureau of Standards, established in 1901, and the National Advisory Committee for Aeronautics, born in 1915. Both have played vital roles in the development and growth of giant new industries and fields of technology.

The Challenge of Today

Today science and technology affect the fabric of modern society in the most pervasive and intensive ways. The powerful focusing of basic and applied science and engineering development which accomplished so much for us in World War II -- and which has had almost unbelievable growth since -- has profound implications for this Nation today and tomorrow. Our national welfare, not only in terms of national survival, but also in terms of our economic growth and social well-being, depends upon a continuation of these concerted scientific-technological efforts. They are of vital importance to maintaining our deterrents to war, to conservation of our natural resources and to the search for new resources to replace those being depleted by our increasing use of them, to the creation of new industries, to improving our capabilities to work together as citizens of a great Nation, and to the pursuit of pure knowledge which is essential to the growth and development of our people.

This situation is recognized by too few of our citizens -- principally only by those already deeply involved in these efforts. We need to become more conscious of this fact as a Nation for two important reasons:

> If our scientific-technological efforts are to continue successfully, there must be continued and increased entry of young people into the scientific and technological career fields.
>
> All our citizens must have a greater knowledge and appreciation of these efforts -- more scientific literacy -- in order to understand and react intelligently to them and help to guide their impact upon us.

Threefold Concern

The Federal Government's concern in this problem is threefold:

> It must continue its role of focusing attention upon proper national objectives in science and encouraging progress toward those objectives. It must assure that no potential field of science is neglected, that gaps are explored and abandoned fields reviewed. This needs to be so because in the search for scientific truths:
>
> educational institutions are necessarily more concerned with their primary goal of education in the sciences.

private industry is necessarily more concerned with those scientific and technological matters which have immediate impact upon its primary goal of earning a profit.

there are certain scientific and technological projects, many of which involve the very limits of knowledge, that must be advanced by the Federal Government because they are beyond the interest or the capabilities of any individual or corporate body. Among these are:

The study of mankind

The development of new forms of energy -- atomic power and beyond

The exploration of the planet earth, its lands, its waters, its resources

The exploration of space.

The Federal Government is the largest employer, both directly and indirectly, of scientists and engineers in the free world. More than 24,000 Federal employees are at work at the professional level in the biological sciences, 30,000 in the physical sciences, 43,000 in medicine, and 90,000 in engineering. And thousands of persons employed by private industry work full-time or part-time on professional assignments related wholly to Government-financed research and development.

The Federal Government expends fifteen billion dollars annually for Federally sponsored research and development performed directly and under contract. Through its programs, the Government now supports over two-thirds of the research and development of the Nation.

CONCEPTION AND DESIGN

The designer must conceive the ideas to carry out the objectives expressed above. The unclassified scientific and technological activities available for inspiring ideas will be those of the Departments of Army, Navy, Air Force, Interior, Agriculture, Commerce, and Health, Education, and Welfare, and the National Science Foundation, the National Aeronautics and Space Administration, the Atomic Energy Commission, and the Smithsonian Institution. No individual agency will be identified in the sense that each will have a separate section of the exhibit. The idea is to present a cohesive, meaningful picture of the World of Federal Science and Engineering rather than stringing together a competing series of agency exhibits.

THE EXHIBIT

The exhibit should:

Dramatically illustrate the broad scope of Federal science and engineering and show that our scientists and engineers are impacting the frontiers of knowledge.

Impress the viewer with the interrelationship of basic science, applied science and technology, and their impact upon man, and dramatically illustrate the fact that these are the tools which enable man not only to adapt to and survive in his environment but to change his environment.

Reveal some of the major areas in which Federal research is seeking new breakthroughs and, where applicable, progress is being made or steps are being taken toward accomplishment of aims.

Explain the growing need in research and engineering work for young people with keen minds and fresh approaches.

The exhibit should be balanced. Because the scientific and engineering efforts in which the Federal Government is involved are so vast in scope, for purposes of their presentation we have categorized Federal efforts along five main themes. These are:

Man and the Basics (Food, Clothing, Shelter)

Man, the Living Being

Man and His Earth

Man and the Universe

Man and the Group

These themes are not restrictive in the sense that a particular Federal project or discovery contributes only to one of them. As a matter of fact, it is extremely doubtful that any Federal project exists which does not, at least through the by-products of knowledge, techniques, or new instruments or tools, contribute to most of the five themes. We note that there is interaction and interdependence among these themes in order to emphasize that, while we list them in order to illustrate the Federal activities in Science and Engineering, we do not require that they be kept separate in the design.

Man and the Basics (Food, Shelter, Clothing) -- Within this theme we are concerned with activities such as man's efforts to improve and increase his sources of food, to preserve his food, to combat diseases and insect pests that destroy his food and fiber. We are also concerned with his efforts to shelter himself, to adapt to unusual living conditions, and to clothe himself. We are also concerned with his search for new forms of energy and his efforts to reduce the drudgery of life and consequently provide greater opportunity for him to enjoy an additional number of spiritual and other cultural pursuits. Examples of Federal activities and/or programs which contribute to this theme are:

> Federal scientists are waging a relentless battle against insects whose depredations on foods, plants, and livestock total billions of dollars each year. Ten thousand harmful species of insects in the United States are arrayed against the scientist. Scientific weapons include insecticides, radiation sterilization of laboratory-reared males, systemic insecticides fed to animals and plants, insect diseases.
>
> Scientists are unlocking a vast storehouse of knowledge about plants by using radioactive isotope tracers to follow the process by which plants take up the materials of the earth to yield fruits, grains, and fibers. They are revamping our poultry and livestock to provide tastier and leaner meats, turkeys designed to fit family needs, chickens that lay better, and cows that give more milk.
>
> Conversion of saline water to fresh is progressing with large-scale demonstration plants constructed and combination nuclear power-desalination plants being planned.
>
> A fish protein concentrate has been developed which can feed 600 million persons a day at a cost of one and one-fourth cents each.
>
> The application of air foam plaster to the construction of military shelters and eventually to housing is progressing.
>
> Federal scientists have developed length-measuring blocks which are accurate to one-millionth of one inch, and weights so delicate that a fingerprint on one of them would destroy its value completely.
>
> A complete city under ice has been developed in Greenland.

Man, the Living Being -- Within this theme we are concerned with activities such as man's efforts to protect his life, to prolong it, and to learn what life is. We are concerned with man's efforts to extend the capability of his five senses through scientific and technological means, with improvement in his mobility and ability to communicate, and with his efforts to adapt to his environment and to control it to some extent. Examples of Federal activities and/or programs which contribute to this theme are:

> Weather modification is one of the greatest challenges to modern science. A program of research and evolution is being carried on to alter the weather in a controlled fashion so that rain falls where there are droughts, rain clouds are dissipated where there are storms, and hail and lightning damages are prevented.
>
> Viruses lie at the threshold of life. They are parasites that cannot multiply in any environment except living tissue. In one experiment, virus particles were isolated as crystals which seemed to have no more life than a lump of coal. When these crystals were rubbed into the leaves of a host plant, they sprang to life again. These are the organisms that are responsible for a variety of cancers in animals. The question Federal scientists are seeking to answer is: Do they cause cancer in man?
>
> Nuclear energy is playing a vital role in the life of our citizens today. It will increasingly affect all the peoples of the earth. It is powering aircraft carriers, nuclear submarines, and small compact devices (systems for nuclear auxiliary power) to supply power for a variety of space and terrestrial uses.
>
> Medical research is making progress in the use of artificial limbs and the replacement of vital organs with artificial ones. Among developments of the latter are the heart pump, an artificial artery, windpipes and heart valves, and an artificial kidney.
>
> In the field of engineering psychology is a concept known as the Man Amplifier. This program may soon reach the stage of practical application. It envisages a servo-powered exoskeleton structure to be worn by a man which will respond to his normal movement and at the same time provide the possibility of amplifying such movement by several orders of magnitude. Most tasks envisioned for the Man Amplifier involve lifting and moving objects which are too heavy for the unaided man.
>
> One of the most challenging fields of development in the Nation today is the research and development of the Supersonic Transport (SST). This development cuts across all departments of the Government.

Man and His Earth -- Within this theme we are concerned with activities such as man's efforts to learn about himself as an inhabitant of earth, to understand his planet, to learn about earth's other creatures, to study natural conditions and wild populations, etc. Examples of Federal activities and/or programs which contribute to this theme are:

> Development of deep-drilling techniques is making possible an attempt to realize an old dream of scientific exploration of the earth's interior. The objective of project Mohole is to drill through the earth's crust into the mantle. From this project scientists hope to learn more about the structure and composition of our planet, its age and origin, the origin and evolution of life through study of the fossils found in the sedimentary layers, and the age and structure of the ocean basins.
>
> Although increasing amounts of time are being devoted to exploration of the sea, oceanography has just begun. The exploration of the sea offers the possibility of support for our increasing population as well as a means of helping to unlock many of the mysteries of our planet and of man's past. The International Indian Ocean Expedition is, for example, the multinational effort to explore scientifically the world's least known ocean. Merely learning more about this ocean's potentially rich and unharvested food resources might make it possible for nations rimming the Indian Ocean to better feed their people and promote their economic development.
>
> Flood prevention, erosion control and irrigation to save the land are constant concerns of Federal scientists and engineers. The benefits of these activities are reflected in terms of millions of square miles of rescued lands.
>
> With our eyes lifted to the planets and outer space, we often forget that a world is still being discovered under our feet. The investigation and charting of the earth is providing great practical benefits. All air and sea transportation depends on maps and charts. Water and air safety are greatly enhanced by the discovery and charting of hazards. Geodetic surveys, precisely tying together the location and elevation of places throughout our country, are basic to large engineering projects -- highways, bridges, tunnels -- and to setting the boundaries of both public and private properties.

Man and the Universe -- Within this theme we are concerned with activities such as man's efforts to understand and explore the universe. Examples of Federal activities and/or programs which contribute to this theme are:

> Space probes, satellites, and manned spacecraft are bringing us closer to an understanding of our universe. Nuclear energy will make interplanetary travel practical, interstellar exploration feasible.
>
> Radio wavelengths, 10,000 times longer than optical wavelengths, and the wide spectrum of observable radio wavelengths greatly extend the possible observation of the heavens in character and in range. Radio astronomy observatories are now receiving and analyzing radio signals from space -- from the sun, the planets, the radio stars, and other sources.
>
> Balloon-borne telescopes, floating about 80,000 feet above the earth, yet pointed and focused from the ground, are being used to study Mars, Jupiter, and certain red giant stars.

Man and the Group -- Within this theme we are concerned with man's relationships with other individuals of his family and community groups, and with the larger social and economic organizations within which he must live and act. Examples of Federal activities and/or programs which contribute to this theme are:

> The effects of urban living on individuals and family groups.
>
> Manpower development and training for workers displaced by technological change.
>
> Research in the processes of learning.
>
> Studies to overcome the causes of poverty.
>
> Interpersonal relationships and mental health.

Scope of Design Contract

The designer selected to execute the exhibit in its final form will be assigned responsibility for preparation of detailed plans and specifications of all structures, illustrative matter, materials, equipment, assembly and disassembly instructions, and other items forming the finished exhibit, subject to final approval by the Civil

Service Commission Exhibit Work Group. Additionally, he will have the responsibility of preparing -- on the basis of these plans and specifications -- a detailed estimate of fabrication costs which can be employed as a "yardstick" before fabrication bids are invited to estimate accurately the probable cost of the exhibit, and after bids are received to decide wisely whether bids received are within reason and whether any low bids seem so low as to have been prepared imprudently. Wherever the practical possibility exists for substituting less expensive materials, equipment, or fabrication methods without serious effect on the overall quality of the exhibit, this possible substitution will be noted in the plans and specifications as a biddable alternative, and the designer's estimate will include a comparison of probable cost of each alternative. Finally, the designer will be assigned responsibility for assuring that the exhibit fabricator conforms exactly to all requirements of the plans and specifications, and that fabrication proceeds on a schedule certain to meet contract deadlines for inspection, final approval, and delivery. This he will accomplish by a series of inspections made jointly with a representative of the Civil Service Commission Exhibit Work Group -- on the premises of the contract fabricator -- of all materials and equipment, finishes, illustrative matter, type faces, and other items making up the exhibit -- before, during, and after fabrication. It will be the responsibility of the designer to bring any and all discrepancies revealed during these inspections to the immediate attention of the Civil Service Exhibit Work Group and its Contracting Officer, and to submit monthly progress reports detailing the status of all parts to be fabricated. Additionally, it shall be the responsibility of the designer to prepare a preventive maintenance schedule based upon expected usage and manufacturer's recommendations.

Construction

Construction contracts to fabricate the selected design will be handled separately through normal Government bid procedures, not only on amount bid but also on contractor's experience, reputation, facilities, and staff.

Display Dates

The exhibit is scheduled for initial display at the Smithsonian's Museum of History and Technology from March 15 to May 20 in the spring of 1965. Following this two-month period it will be displayed, preferably in full, at other large museums around the country. It will then be refurbished and displayed at the Museum of History and Technology for two months in the spring of 1966.

Exhibit Area

The exhibit will be displayed first in the special exhibits area of the Museum of History and Technology. This area is rectangular in shape, measuring approximately 100 by 51 feet for a total of about 5,000 square feet. The number of persons who might view the exhibit in a single day is estimated to range from 10,000 to 40,000.

Exhibit Devices

With the exception of space, funds, and good taste, there are no restrictions on exhibit design or exhibit devices that may be used. The exhibit must be designed in accordance with standard exhibit measurements to permit easy disassembling, shipping, and erection in other locations. *In addition, a central core of approximately 1,200 square feet must tell a unified story so that part of the exhibit can be shipped to those museums which cannot accommodate an exhibit larger than 1,200 square feet.*

Public Information Office
U. S. Civil Service Commission
August 11, 1964

APPENDIX B

APPENDIX B

UNITED STATES CIVIL SERVICE COMMISSION
Public Information Office
Washington, D. C.

FACT SHEET ON

THE FEDERAL SCIENCE AND ENGINEERING EXHIBIT --

THE VISION OF MAN

The Vision of Man -- The Federal Science and Engineering Exhibit -- portrays the productive partnership of science and Government from the Federally funded Lewis and Clark Expedition in 1803 to today's multi-billion-dollar annual appropriation for research and development programs. The exhibit will be on public display in the Special Exhibits Hall of the Museum of History and Technology of the Smithsonian Institution from April 7 through May 15, 1965, following an official opening ceremony on the evening of April 6.

The 5,000-square-foot exhibit represents the first interagency effort to present a single cooperative display reflecting the Government's broad involvement and interest in science and technology. In sketching the story of Federal conduct and support of scientific and engineering endeavors and their contribution to the Nation's development, security, and welfare, the exhibit seeks to stimulate study and to arouse interest in careers in science and engineering. Since Federally conducted and supported research and development projects are so many and varied, the exhibit provides a sampling of significant and dramatic activities, accomplishments, and objectives.

PROJECT BACKGROUND

In its report to the President on "The Competition for Quality," the Federal Council for Science and Technology urged that the Civil Service Commission "provide leadership and assistance, in cooperation with agencies and departments concerned, to communicate to scientists and engineers as potential employees the opportunities and professional challenges offered by science in the public service, and to inform the general public more explicitly of the scope and achievements of Government science and technology."

The idea of a traveling interagency exhibit that would open in Washington and subsequently tour major American cities was first explored with the Smithsonian Institution in 1962, and member agencies of the Federal Council for Science and Technology were invited to participate in a joint effort to develop such an exhibit. The story line and plans for interagency cooperation in finding and developing the display were worked out during 1963 and 1964 by a work group of public information and exhibit specialists from the participating agencies, under leadership of the Civil Service Commission's Public Information Office. In the spring of 1964, the heads of the participating departments and agencies accepted the exhibit project plan advanced by OSC Chairman John W. Macy, Jr.

Leading industrial design firms were invited to submit conceptual presentations for the exhibit in September 1964, and the firm of Herbst-LaZar of Chicago was selected to design the display from among nearly a score of firms that responded. Competitive bids for fabrication of the display were announced in December 1964, and General Exhibits, Inc., of Philadelphia was awarded the construction contract.

Agencies cooperating in developing the display were the Departments of Army, Navy, Air Force, Interior, Agriculture, Commerce, and Health, Education, and Welfare, the Atomic Energy Commission, the National Science Foundation, the National Aeronautics and Space Administration, and the Smithsonian Institution. The project was coordinated by the Civil Service Commission.

EXHIBIT OBJECTIVES

Primary objective of the exhibit is to stimulate student interest in the study of science and engineering to assure that the Nation's ever increasing demands for top talent in these fields will be met in the years ahead. Government, as the country's largest user of skills in these areas, will be among the employers benefiting. It is hoped that the exhibit will contribute to the development of "scientific literacy" in our society, which leading scientists and statesmen have called a critical need because of the socioeconomic changes stemming from the knowledge explosion and technological revolution of the mid-20th century. And it is expected that the display will improve public knowledge of and regard for the work and achievements of Federal and other scientists and engineers.

THEME AND CONTENT

The "Vision of Man" theme evolved from the story-line concept that the display depict man's constant quest for knowledge, the interrelationships of scientific disciplines, and the impact of science on man. The theme is from Alfred Lord Tennyson's poem, "Locksley Hall," which contains these lines:

For I dipt into the future,
  far as human eye could see,

Saw the Vision of the world,
  and all the wonder that would be.

The exhibit is divided into the following sections: Introduction, Man, the Living Being, Man and the Basics, Man and the Universe, Man and His Earth, Man and the Group, and a Career Guidance Center. However, the core of the display has been designed as a unified whole so that the interrelationships of the sciences and their effect on man are felt. For example, an earth symbol is the centerpiece of the display, but by the placement of transparent panels and use of lighting effects the viewer standing in the Living Being Section may look from one corner of the exhibit through the Earth Section and beyond to the opposite corner and see the environment of space suggested in the Universe Section. Or amid the exhibit features describing man's exploration of the outer reaches of the universe, he may glance in the opposite direction to see the DNA model representing scientists' search for the secrets of the hereditary code confined in the molecule.

Descriptions of the separate sections and their objectives follow.

Introduction

The introductory area illustrates the historic and creative partnership between science and the Federal Government and provides a preview for each of the areas that follow. It features cases containing graphics, models, and artifacts that pay tribute to some of the early explorers, inventors, and scientists who received Government support in their outstanding endeavors. Among the items of interest are: Lewis and Clark's compass, a model of the Wright Brothers' plane, Lawrence's first cyclotron, Goddard's first rocket, and an early plow; artifacts from the Mesa Verde excavations; the skeleton of a prehistoric golden eagle; and the Krypton-86 Lamp.

Man, the Living Being

This section spotlights man's efforts to protect and prolong life -- and to ascertain what it is. The focus of attention is an animated DNA (deoxyribonucleic acid) model used to explain scientists' efforts to decipher the coded language of heredity locked in the DNA molecule. Also illustrated in the area are -- research projects to amplify and extend man's mental and physical abilities, to prepare man for long space journeys, to help him fly at supersonic speeds, to learn to adapt to undersea environment, and to augment, or perhaps replace, failing or defective organs. Items of interest include a man-amplifier model (pedipulator), fluid-amplification respirator, variable-wing plane model, heart pacer, and an illustration of research in skin communication.

Man and the Basics

This section illustrates man's continuing efforts to improve and increase his sources of food, to preserve his food, to combat diseases and insect pests; his efforts to protect, clothe, and shelter himself; his search for new forms and sources of energy and efforts to reduce drudgery. Points of interest include models of water desalting plants, demonstration of microwave transmission of energy, new products (novawood and stretch cotton), the story of phytochrome (light control of plant growth) and the eradication of the screwworm fly (by irradiation-sterilization of the fly in the pupal stage), and illustration of use of irradiation in plant genetics and food preservation.

Man and the Universe

This section illustrates man's efforts to explore and understand the universe. It shows how he is using earth-based laboratories, balloon- and rocket-borne instruments, satellites, manned spacecraft, and optical and electronic instruments to probe near and far space to answer age-old questions about the universe. It spotlights his development of better propulsion systems, his study of the atom and its particles, his inquiry into the nature of matter, the promise of atomic fusion, his designing of new spacecraft and systems to support men on deep space journeys, and his studies to help man adapt to the environment of space. Points of interest include models of a radio telescope, nuclear rocket engine (NERVA), and space laboratories -- MOLAB and GULLIVER -- representative of the types of research vehicles to be landed on the moon and planets; a MARINER IV model and progress track; and a space-tool demonstration device.

Man and the Earth

This section illustrates man's continuing efforts to learn more about his earth and other living creatures on it, and to find, benefit, and conserve the resources of his planet. It shows how he is learning about the earth's origin, age, and structure; illustrates how he is analyzing earth's atmosphere with airborne instruments; spotlights his efforts to find out how birds and fish migrate; and tells of his study of the chemistry of the sea and research to garner more of its riches. Interest points include models of undersea research craft, MOHOLE cores and drills, and recordings of fish sounds.

Man and the Group

This section focuses on man's relationships with his fellow men in an era of accelerated social, economic, and technological change. It underscores the fact that if the advances in science and technology are to have meaning, man must learn more about himself in his social setting. It accents the importance of research to bring better understanding of man as a member of his group and to help mankind achieve its highest purposes. Dramatic pictures and recorded statements point up some areas of central concern -- education, health, housing, old age, and conservation -- and illustrate how the partnership of science and Government is seeking to solve problems of man in our increasingly interdependent society.

Career Guidance Center

The Career Guidance Center will feature a panel on the increasing need for scientists, engineers, and technicians in the future and about the opportunities for scholarships, fellowships, and grants available for study in these fields. Attendants in the Guidance Center will be available to answer inquiries and to suggest sources for additional information about career opportunities and assistance for science and engineering study.

RELATED FILM SHOWINGS

A presentation of selected color films, showing various aspects of the participating agencies' work in science, will take place in the auditorium adjacent to the Exhibit Hall. Films will be shown between 11 a.m. and 12 noon and between 2 and 3 p.m. on Mondays, Wednesdays, and Fridays from April 7 to May 14, and each film will be repeated two or more times during the period of the Exhibit. A list of the film titles, giving the exact times when each is to be shown, will be available at the auditorium.

EXHIBIT TOUR SCHEDULE

Following its premiere in the Special Exhibits Hall of the Museum of History and Technology of the Smithsonian Institution from April 7 through May 15, 1965, the Vision of Man will be featured in the Federal Pavilion at the New York World's Fair from May 20 until the Fair closes in October 1965. It will then be displayed in the California Museum of Science and Industry in Los Angeles and in the Museum of Science and Industry in Chicago in the fall of 1965 and winter of 1966 before returning to the Smithsonian Institution in the spring of 1966.

SCIENCE AND GOVERNMENT

The Federal Government has been concerned with science and technology since the beginning of the United States of America. Article I, Section 8, of the Constitution provides: "The Congress shall have the power ... to promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries." But this rather inconspicuous nod to science in the Constitution does not reveal the extent to which the idea of the Federal Government's encouragement of science loomed in the minds of many of our Nation's founders.

At the time of the founding of our Nation, the relationship of science and technology and the impact of one upon the other was not really demonstrable. Generally, inventions were the result of the ingenuity of craftsmen working on a trial and error basis rather than the result of careful application of scientific theory. However, the usefulness of science had gained acceptance. Physics and astronomy were proving of value to land surveyors, ships' captains, and military men, and natural history had a close alliance with medicine. But the body of scientific knowledge had not grown so great nor so specialized that its limits were beyond the reach of educated persons. Knowledge of natural history and natural philosophy was part of their culture.

As educated men, the framers of the Constitution were familiar with science and valued it. (Benjamin Franklin, for example, was one of the greatest scientific minds of his time.) Their knowledge and appreciation of science and a belief that the new Nation had a responsibility to diffuse knowledge among its people led them to give consideration to the constitutional position of science in the government they were trying to form.

As a result of Article I, Section 8, of the Constitution, the first patent law, designed to encourage individual ingenuity and secure for inventors some benefits of their creativity, was enacted in 1790; the Patent Office itself goes back to 1802. In the same year, engineering was given emphasis by the Federal Government when President Jefferson created a Corps of Engineers which "shall be stationed at West Point in the State of New York and shall constitute a Military Academy."

The first actual scientific activity of the Government dates to 1807 when the Congress authorized a survey of the coast and established the Government's first technical bureau -- existing today as the Coast and Geodetic Survey. Earlier governmental involvement in scientific activity took the form of sponsorship of ventures such as the Lewis and Clark expedition in 1803 -- which made significant findings in botany and zoology -- and S.F.B. Morse's testing of the telegraph. Other landmark actions included the establishment of an Office of Weights and Measures (forerunner of today's National Bureau of Standards), the legislative requirement that the Patent Office test each invention (calling implicitly for the use of scientific principles), and the creation of the Naval Observatory -- all of which came about before the mid-19th century. Another significant early development was the establishment of the Smithsonian Institution in 1846, under terms of the will of James Smithson.

As new departments and agencies were established in later years, their enabling legislation began to include specific provision for a scientific function, sometimes reflecting congressional intent that the department or agency become the repository of the most authoritative information related to the missions of the organization. The classic example is the 1862 act creating the Department of Agriculture, which requires the agency "to acquire and diffuse ... useful information on subjects connected with agriculture in the most general and comprehensive sense of that word, and to procure, propagate, and distribute among the people new and valuable seeds and plants."

The Department of Agriculture's creation marked the beginning of the era of permanent bureaus with scientific functions. The next 48 years saw the establishment of a growing number of such organizations, among them the Signal Corps, the Naval Hydrographic Office, the Fish Commission, the U. S. Geological Survey, and the Weather Bureau.

With the 20th century and the emergence of large-scale industry as the dominant force in the Nation, a new type of Government scientific organization came into being, in response to the developing close relationship between industry and science. Noteworthy among these organizations were the National Bureau of Standards, established in 1901, and the National Advisory Committee for Aeronautics, born in 1915. Both have played vital roles in the development and growth of giant new industries and fields of technology.

More recently, the increasing importance of science and technology in our society has been reflected in the establishment of such agencies as the Atomic Energy Commission in 1946, the National Science Foundation in 1950, and the Federal Aviation Agency and the National Aeronautics and Space Administration in 1958.

Today science and technology affect the fabric of modern society in the most pervasive and intensive ways. The powerful focusing of basic and applied science and engineering development which accomplished so much for us in World War II -- and which has had almost unbelievable growth since -- has profound implications for this Nation today and tomorrow. Our national welfare, not only in terms of national survival, but also in terms of our economic growth and social well-being, depends upon a continuation of these concerted scientific-technological efforts. They are of vital importance to maintaining our deterrents to war, to conserving our natural resources and finding new resources to replace those being depleted by our increasing use of them, to the creation of new industries, to improving our capabilities to work together as citizens of a great Nation, and to the pursuit of pure knowledge which is essential to the growth and development of our people.

If our scientific-technological efforts are to continue successfully, there must be continued and increased entry of young people into careers in science and engineering. And the citizens of the United States should have a greater knowledge and appreciation of these efforts -- more scientific literacy -- in order to understand and react intelligently to them and help to guide their impact upon us.

March 25, 1965

APPENDIX C

PHYSICAL LAYOUT OF LOS ANGELES EXHIBIT

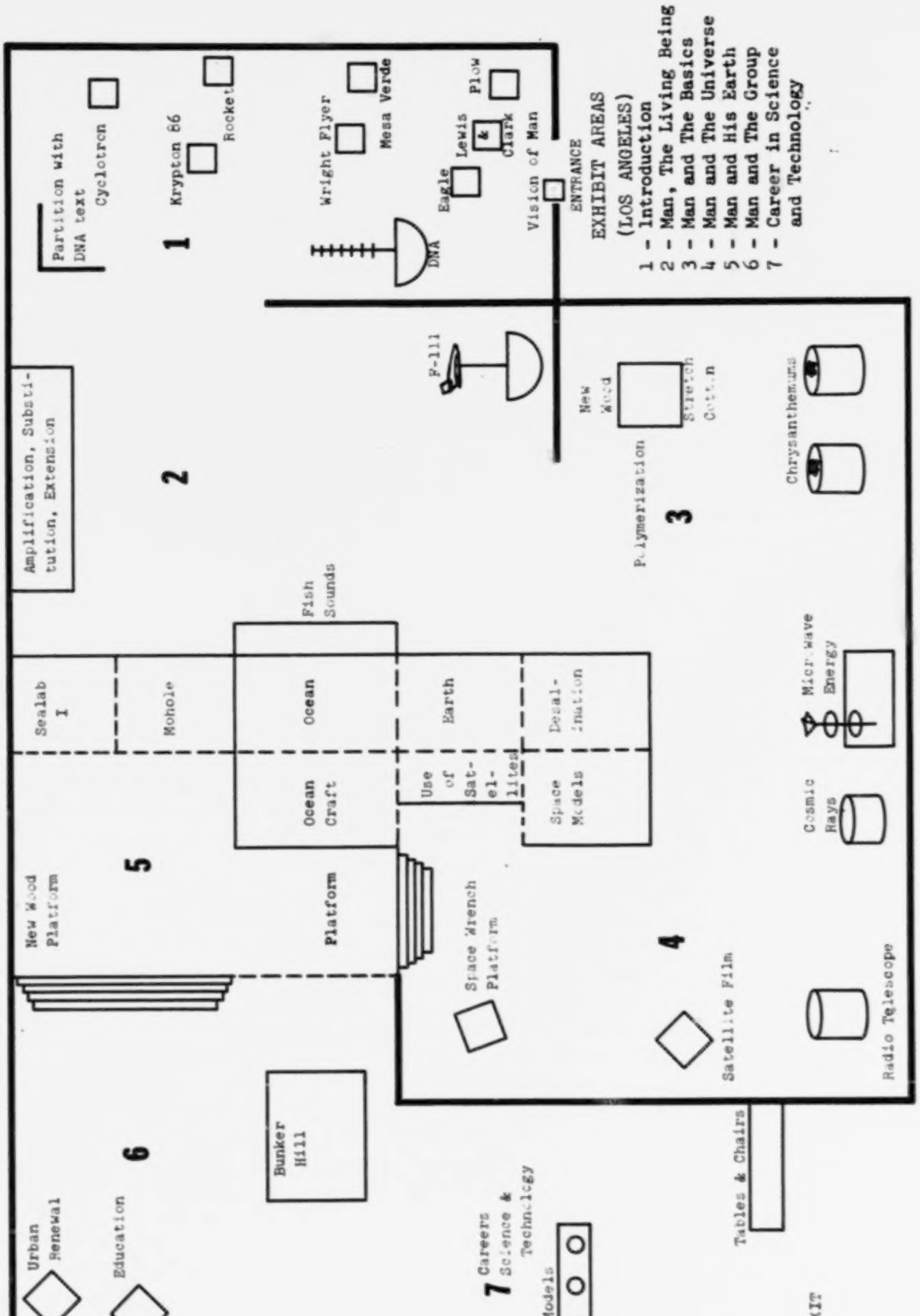

Partition with DNA text
Cyclotron
Krypton 86
Rocket
Wright Flyer
Mesa Verde
Lewis & Clark
Plow
Eagle
Vision of Man
ENTRANCE
DNA
1
F-111
Amplification, Substitution, Extension
2
New Wood
Stretch Cotton
Polymerization
3
Chrysanthemums
Fish Sounds
Sealab I
Mohole
Ocean
Earth
Desalination
Ocean Craft
Use of Satellites
Space Models
Micro Wave Energy
Cosmic Rays
5
New Wood Platform
Platform
Space Wrench Platform
4
Satellite Film
Radio Telescope
Bunker Hill
6
Tables & Chairs
Urban Renewal
Education
7
Careers Science & Technology
Models
EXIT

EXHIBIT AREAS
(LOS ANGELES)

1 - Introduction
2 - Man, The Living Being
3 - Man and The Basics
4 - Man and The Universe
5 - Man and His Earth
6 - Man and The Group
7 - Career in Science and Technology

APPENDIX D

POSTTEST BACKGROUND QUESTIONS

INSTRUCTIONS: Please check or fill in each of the items below. If an item does not apply to you, simply skip to the next item.

1. What is your age? _______
2. Check one: _____ Male _____ Female
3. Are you currently a full-time student in: ___ high school
   ___ college
4. If you are currently employed full time, what is your occupation and job title? ______________________________
5. If you are not currently employed full time, specify your father's occupation and job title. (Do likewise for your mother, if she is employed.)
6. Where do you live? ___ urban area ___ suburban ___ rural
7. What is the highest educational level you have achieved?
   Elementary - 6 7 8 High School - 1 2 3 4
   College - 1 2 3 4 5 6 7 8
8. How many science courses did you take in high school? (You may not remember exactly, but try to estimate the number.)
9. What science subjects did you study in high school?
10. How many science courses did you take in college (including any courses in graduate school)? __________
11. What science subjects did you study in college (including any courses in graduate school)?
12. How would you rate your interest in science in comparison with your interest in other fields? Not interested, a little interest, a moderate (fair) amount of interest, a lot of interest, extremely interested
13. How long did you spend looking at the exhibit? ___________
14. Did you come to the museum especially to see this exhibit?
15. What did you know about this exhibit before you came?
16. How did you hear about this exhibit, for example, through newspapers, TV, friends, etc.?
17. Are you interested in studying science? ___ Yes ___ No
18. Do you currently have a scientific career or an interest in obtaining such a career? ___ Yes ___ No

APPENDIX E
POST-INTEREST INDEX

1. Did this exhibit stimulate your interest in a particular topic or area? ____ Yes ____ No

   If you answered yes, what topic(s) or area(s)?

   ____

   ____

   ____

2. Rank the exhibit areas below in order of their interest to you, as well as you can from the description given. Put a 1 next to the area you like best, 2 for the next best, etc., and 5 for the area you like least.

   _____ a. studies to learn more about the earth and living creatures, and to learn how to conserve the earth's resources.

   _____ b. studies to improve man's relationships with his fellow man.

   _____ c. studies to find ways to improve food, clothing, and shelter, and develop new forms and sources of energy.

   _____ d. studies to find ways to protect and prolong life and to determine what it is.

   _____ e. studies designed to explore space.

APPENDIX F

APPENDIX F

ATTITUDE QUESTIONS

INSTRUCTIONS: Please circle or check one of the answers for each of the questions below (except where you are given special instructions).

1. To what extent do you think science has influenced you?
   not noticeably, somewhat, quite a bit, significantly

2. If your son (or daughter) were planning a career in science, would you prefer to have him work, after graduation, for: the Federal Government, a university, private industry, a nonprofit research laboratory

3. In your opinion, most projects designed to obtain basic scientific knowledge should be financially supported (but not actually conducted) by: industry, private research laboratories, the Federal Government, universities

4. When you are reading newspapers and magazines of various kinds, do you read articles on scientific research and the applications of scientific research: rarely, if ever; seldom; usually; almost always

5. In your opinion, the majority of projects that make significant contributions to scientific knowledge require: the efforts of scientists from only one scientific field, at least some interaction between scientists from different fields, a fair amount of interaction between scientists from different fields, a great amount of interaction between scientists from different fields.

6. The equipment, facilities, and staffing of the Federal Government's scientific research laboratories are: inadequate, comparable to those found in private laboratories, generally not as adequate as those found in private laboratories, generally better than those found in private laboratories

7. In your opinion, most projects designed to obtain basic scientific knowledge should actually be conducted by: universities, private research laboratories, the Federal Government, industry

8. To what extent do you think our nation's economic prosperity depends on our scientific research and development? not noticeably, somewhat, quite a bit, significantly

9. In your opinion, most projects that use basic scientific knowledge to develop useful, practical things should actually be conducted by: the Federal Government, industry, universities, private research organizations

10. If you had detailed control over how your tax dollar was spent, which of the following projects would you prefer to help support? (Check as many as you want.) building hospitals and mental health institutions, investigating resources on the ocean floor and below the earth's crust, investigating ways to influence heredity, building better education facilities, building superhighways, investigating the effectiveness of new educational techniques

11. If you checked more than one of the projects in the above question, check below the project that you would support if you could contribute to only one project: building hospitals and mental health institutions, investigating resources on the ocean floor and below the earth's crust, investigating ways to influence heredity, building better education facilities, building superhighways, investigating the effectiveness of new educational techniques

12. Do you think the Federal Government should: not support scientific research, concentrate its support on applied scientific research, concentrate its support on basic scientific research, distribute its support over both basic and applied scientific research

13. Do you think that scientific improvement of man's intelligence and other human characteristics is: not very likely, fairly likely, quite likely, almost certain

14. In your opinion, most projects that use basic scientific knowledge to develop useful, practical things should be financially supported (but not actually conducted) by: universities, the Federal Government, industry, private research laboratories

15. The amount of my tax money being spent on scientific research is: too much, about right, not enough

16. Which of the following specialists are employed by the Federal Government? (You may check as many as you want.) lawyer, anthropologist, pharmacist, astronomer, biochemist, forest ranger, mathematician, engineer, sociologist, fishery specialist, geographer, geodesist, geneticist, meteorologist, psychologist, speech pathologist, oceanographer, physicist, plant scientist, technical writer

17. In the last twenty years do you think the increase in scientific knowledge and capabilities has been: barely noticeable, small, fairly high, very high

18. If you had free choice, in which one of these fields would you like to continue your education: a physical science, for example, physics or chemistry; business management; law; a biological science, for example, botany or zoology; engineering; a social science, for example, anthropology or psychology; music; journalism

19. To what extent do you think practical applications of science have influenced you? not noticeably, somewhat, quite a bit, significantly

APPENDIX G

OPEN-END CONCEPT QUESTIONS

INSTRUCTIONS: Please answer each of the items below as briefly as possible on the lines provided below the question.

1. How do porpoises locate objects?
2. What is Project Mohole?
3. What is the significance of designing aircraft with variable wings?
4. What is phytochrome?
5. What is the importance of the development of the radio telescope?
6. How does the current concept of the atom differ from the Greek concept?
7. What is the significance of DNA?
8. What are cosmic rays, and where do they come from?
9. What problem made it necessary to design a special wrench for use in outer space?
10. What is Molab?
11. What is the purpose of the Sealab experiment?
12. What is the function of a microwave transmitter?
13. The world faces a water shortage. What is the most promising approach now being worked on to solve this problem?
14. Why will two kinds of rockets probably be used for deep-space missions?

INSTRUCTIONS: If you know several answers to any of the questions below and you have already seen the exhibit, give the answer based on information provided in the exhibit.

1. Our nation's greatest resource is our people. Name as many areas as you can in which the Federal Government is helping to make the best use of this resource.
2. What is "novawood"?
3. In the exhibit on Man and the Basics, a model of an old colony strong plow was shown with a picture of Abraham Lincoln. Describe, in your own words, the relationship between Lincoln and the plow.

APPENDIX H

APPENDIX H

OPEN-END KNOWLEDGE QUESTIONS

INSTRUCTIONS: Please answer each of the items below as briefly as possible on the line provided below the question.

① What process has been found to speed up the mutation rate of living cells?

2. Identify the nation's largest employer of scientists, engineers, and technicians.

③ Man has developed a technique for locating hidden objects that is very similar to the technique used by porpoises. What is this technique called?

4. What is an accelerator used to bombard?

5. Label the four layers/divisions of the earth on the drawing below.

6. Scientists have recently discovered additional supplies of minerals, oil, natural gas, and diamonds. Where have these supplies been discovered?

7. In what kind of nuclear rocket engine are cesium atoms heated?

8. What field of study is concerned with the earth's size, shape, mass, and gravitational field?

⑨ What device has been developed recently which enables astronomers to see beyond interstellar dust clouds?

10. What proportion of the earth's surface is covered by water?

11. What is the name of the chemical process used to create new chemical substances by applying the close-fit principle?

12. Many scientists believe that they have identified the earliest form of DNA. What is it called?

13. What is the name of the project designed to answer such questions as: "How old is the earth?", "What is its origin?", and "Is it getting hotter or colder?"

⑭ What is the name for high energy particles that streak into the earth's atmosphere from outer space?

15. To make ordinary wood stronger and more beautiful, it is soaked in a "bath" and passed through a beam. What kind of "bath" is used, and what is the beam called?

16. What is the name of the vehicle which has a complete life-support system and may be sent to land on the moon?

17. What operational characteristic of aircraft have scientists been able to control by applying the "area rule" concept?

INSTRUCTIONS: If you know several answers to any of the questions below and you have already seen the exhibit, give the answer based on information provided in the exhibit.

1. Name one method of desalinating water.
2. What "partnership" was described by this exhibit?
3. Can useful energy be transmitted from one point to another without using wires? ____ Yes ____ No If so, how?
4. What is one method used to study the migratory habits of commercial and game fish?
5. What was unusual about the wooden platform in the exhibit?
6. What is the name of this exhibit?
7. What is the estimated increase in total school enrollment over the period 1964-1974?

⑧ What is a recent and important scientific discovery that promises to give man more control over plant growth?

9. What term is used to describe the translation of basic scientific knowledge into useful, practical things?

⑩ What scientific technique have scientists developed for getting rid of insects?

⑪ What are the names of two theories of the universe?

12. What kind of light beam can be used to destroy a tumor on the retina of an eye?
13. Who paid for this exhibit?

APPENDIX I

APPENDIX I
MULTIPLE-CHOICE QUESTIONS

INSTRUCTIONS: Please check or circle an answer for each of the questions below. Answer all of the questions in the order they are given; do not skip around. Even if you are not sure of an answer, take a guess. There is no time limit on this test. Some of these questions may appear to have more than one right answer. However, if you have seen the exhibit, you should select the answer based on the information provided in the exhibit. Remember, answer every question.

1. The project called Sealab 1: lasted three weeks, was an experiment in undersea mining, was an experiment in undersea living, was the first stage of Project Mohole.
2. Fish sounds may be converted to a visual pattern by an instrument called the: spectroscope, radio telescope, oscilloscope, opthalmoscope.
3. In the United States, the first major step toward a scientific approach to agricultural problems was taken when: the National Grange was established by the farmers, the old colony plow was improved, the screwworm fly was eradicated, land-grant agricultural colleges were established by the Government.
4. In designing the F-111, engineers primarily made use of knowledge gained from: the development of the ion engine, Goddard's paper on reaching extreme altitudes, experiments with the variable wing, the field of rocket astronomy.
5. Many of the improvements in food, clothing, and housing materials have come about due to the: translation of science into technology; use of scientific principles in everyday life; pursuit, attainment, and arrangement of basic knowledge; translation of technology into science.
6. John Dalton was essentially correct when he said that: a given element may have more than 100 different atoms, the atom is made up of many particles, the atom is indivisible, all atoms of a given element are identical, possessing a unique atomic weight.
7. Eventually, scientists hope to unlock energy from the fusion of heavy: nitrogen atoms, oxygen atoms, hydrogen atoms, cesium atoms.
8. The first satellite to broadcast the human voice from space was: Beacon, Geos, Pageos, Score.
9. Today, the main problem in using the multiple flash technique to convert large amounts of salt water into fresh water is the cost of: fresh water distribution, storing fresh water, the machinery needed, the energy required.
10. An example of basic science is the: construction of the F-111, investigation of subatomic particles, eradication of the screwworm fly, development of novawood (plastic wood).

⑪ The discovery of phytochrome was made by: florists who used it to design ornamental flowers, scientists who worked for the Government, oceanographers who used it in studying fish, chemists who used it in desalination.

12. The floating drilling rig used in Project Mohole will be kept in position by: a 4,000-ton electromagnetic system, a celestial navigation system, an acoustic and electronic system, a microwave energy system.

13. How many distinct kinds of chemical subunits can pair up to form the DNA molecule? 4, 6, 8, 12.

⑭ Scientists have learned that the energies of cosmic rays: are sometimes more powerful than accelerator-generated energies, are usually too weak to pass through a solid material, have often been destructive to property, can be used to guide unmanned rockets.

15. In producing the "new wood," wood is soaked in a bath of: sodium hydroxide and then dried by warm air, monomers and then a beam of radiation is passed through it, guanine and then a beam of radiation is passed through it, potassium chloride and then dried by warm air.

⑯ During the total solar eclipse of 1958, eight NIKE-ASP rockets gathered data to show that: x-rays originate in the gaseous corona, both x-rays and ultraviolet rays originate in the sun's surface, x-rays originate in the sun's surface, both x-rays and ultraviolet rays originate in the gaseous corona.

⑰ It has been discovered that plants grown directly from irradiated seeds: always show desirable changes, mostly show desirable changes with a few undesirable changes, mostly show undesirable changes with a few desirable changes, always show undesirable changes.

18. Enough electricity to light New York City for one hour can be produced by the fission of one pound of: deuterium, uranium, cobalt, tritium.

19. Our Government is providing support in solving problems of: education, aging, urban transportation, all of the above.

⑳ The volume cycle respirator is an example of: modification, extension, substitution, amplification.

㉑ Scientists are investigating the atolla wyvillei to find out: how the fish uses the echo-locate principle in navigation, how the fish makes its chemical light, why the fish migrates long distances to spawn, why the fish lives in warm, coastal waters.

㉒ The radio telescope is adding to science much significant information about the: transmission of radio waves to other parts of the world, four layers of the earth, weather conditions throughout the world, mapping of space bodies in the universe.

23. America's pioneer in liquid-fuel rocketry was: Marshall Nirenberg, R. J. Van de Graaff, Robert H. Goddard, Richard T. Whitcomb.

24. A place near Hawaii was selected for Project Mohole because scientists know that at this location: the mantle is less dense, the Mohorovicic Discontinuity is thinner, the earth's crust is thinner, the earth's central core is closer to the surface.

25. A chemical injected into fish for tracking purposes is: thymine, tetracycline, pion, adenine.

㉖ It has been established that mutations: are apparent only in plants and lower life forms, can be induced only by scientific techniques, are genetic alterations, can usually be seen with the naked eye.

㉗ The subatomic particle which was first found in cosmic rays and is associated with the force that binds the atom's nucleus together is the: pion, positron, ion, baryon.

28. Nerva's engine is powered by: chemical energy, microwave energy, electrical energy, nuclear energy.

29. By the year 2000, it is believed that our vast power requirements will be met by energy generated from: fossil fuels, fissionable fuels, particle accelerator plants, microwave transmitters.

30. The Mariner IV has been called our nation's "most exacting space probe" because it traveled: close enough to the sun to take photographs of the gaseous corona, to Jupiter and sent back photographic data, to Mars and had a soft landing, to Mars and sent back photographic data.

31. The scientific expedition led by Lewis and Clark was significant because it was the first to: use a compass, map the Southwest territory of the United States, return with artifacts from Mesa Verde, be supported by the Federal Government.

32. A unique application of a centrifuge is: to counter the effects of weightlessness in space, in the chemical process of polymerization, in the reverse osmosis process of desalination, to protect the space traveler from radiation.

33. In making wood-plastic flooring and in speeding up the mutation rates of plants, a common technique has been the use of: radiation, polymerization, phytochrome, chemicals.

34. The creation of DNA probably resulted from the formation of: nucleic acid chains in the ocean, amino acid chains in plants, nucleic acid chains in plants, amino acid chains in the ocean.

35. The laser beam is now being used in: creating controlled nuclear fusion, destroying tumors, drilling the first stage of Project Mohole, reproducing DNA from chemical components.

㊱ The study of radio wave sources has resulted in the: discovery of unknown hydrogen gas clouds, discovery of the origin of cosmic rays, development of the tiny shortwave receiver, development of wireless energy.

37. The Trieste is: a multi-purpose aircraft, a deep-sea exploration ship, an experimental spacecraft, a synchrocyclotron.

38. Special zero-reaction tools have been designed to use in space due to the: belts of high-intensity radiation, lack of gravity, presence of meteroids, solar winds in interplanetary space.

39. The transmission of microwave energy is an important research breakthrough because such energy can be used to: perform useful tasks without wiring, speed up molecular changes in polymerization, desalinate seawater at a low cost, create desirable mutations in plants.

40. Cotton fiber is stretched by being immersed in a solution of: monomers, potassium chloride, maleic hydrazide, sodium hydroxide.

41. In tearing apart an atom, an accelerator is used to: speed up electrons, protrons, or deuterons to bombard a nucleus; speed up deuterons to bombard electrons; speed up the protrons of a nucleus to bombard electrons; speed up the nucleus of the atom to make subatomic particles divide and spin away.

㊷ The scientific technique developed to eliminate the screwworm fly was: sterilization, implantation, mutation, poison.

43. The formation of the area rule concept was an important technical breakthrough in: developing the bubble chamber, making new maps of the universe, designing supersonic aircraft, increasing agricultural yield.

㊹ Phytochrome is a: chemical injected into fish for tracking purposes, chemical in seawater that determines a salmon's migration, catalyst used in desalting seawater, coloring matter in plants that influences plant growth.

45. The effect of DNA is determined by: whether a species has DNA molecules, the kinds of chemical units that make up the DNA molecule, the size of the DNA molecules, the sequence of the DNA molecule's chemical units.

46. Project Mohole will: drill only to the Mohorovicic Discontinuity, be completed by 1968, be paid for by the Government, be done in cooperation with Sealab.

47. A new 11-state intertie system which extends from the Pacific Northwest through the Southwest will be built to: better utilize power resources, improve transportation systems, facilitate radio communications, transmit microwave energy.

㊽ The "father of rocket astronomy" is: Richard T. Whitcomb, R. J. Van de Graaff, Ernest O. Lawrence, Herbert Friedman.

49. The significant feature of the spacecraft Gulliver is that it: will have a complete life-support system on board, is designed to collect dust on sticky cords, is designed to land on the moon and then return to earth, will be the first manned spacecraft to land on Mars.

50. High energy physics is the study of: x-rays from the sun, high voltage circuitry, subatomic particles, centrifuges

(51) A fish that migrates from the ocean to fresh water to lay its eggs is the: porpoise, atolla wyvillei, goldfish, salmon.

52. After being energized by an accelerator, subatomic particles leave their tracks behind when they are passed through a bubble chamber containing: liquid sulphuric acid, liquid carbon dioxide, liquid hydrogen, liquid oxygen.

53. The research craft called Sealab 1 was: a long steel tube, a floating instrument platform, an unmanned submarine, powered by nuclear energy.

(54) The pedipulator is an example of: perception, amplification, substitution, extension.

55. The largest employer of scientists and engineers is: universities, nonprofit research institutions, private industry, Federal Government.

56. Tiros and Nimbus have been developed to: eventually land on the moon, explore the sea floor, study undersea life, report weather conditions.

57. Propulsion research has shown that the ion engine will take over for the chemical engine during deep space travel because the ion engine: is high on thrust and low on "miles per gallon," is high on "miles per gallon" and low on thrust, has both very low thrust and "miles per gallon," has both very high thrust and "miles per gallon."

58. The microwave energy transmitter was developed: by the United States Air Force, jointly by the United States Air Force and Raytheon, jointly by the Atomic Energy Commission and General Electric, by RCA.

59. Aside from the magnet itself, Lawrence's first cyclotron was constructed mainly of: metal and sealing wax, glass and sealing wax, wood and glass, metal and plastic.

(60) The distance over which astronomers can detect objects has recently been greatly increased through the use of: photographic plates, optical telescopes, spectroscopes, radio telescopes.

(61) The screwworm fly has been a costly insect because it eventually kills: livestock; crops such as oats; plants found in Florida, Georgia, and Alabama; fruit trees.

62. The Indians who settled in Mesa Verde built their homes in: cliffs, underground tunnels, trees, sand dunes.

63. The scientific research accomplished at the Rancho La Brea tar pits was supported by the: California Museum of Natural History, American Society of Archeologists, Federal Government, American Museum of Natural History.

64. An unmanned life-support system which may be sent to land on the moon is called: Mariner IV, Nerva, Molab, Gulliver.

65. One of the orange-red lines in the Krypton spectrum provides the international standard for measuring: length, saturation, hue, brightness.

66. The pacemaker is used to correct irregularities due to: eye diseases, heart diseases, plant diseases, lung diseases.

67. In the chemical process of polymerization, molecules are: activated to provide thrust for spacecraft takeoffs, collected at positive and negative electrodes, separated temporarily to provide the stretch in cotton, streamlined into more closely fitting units to provide a stronger product.

(68) Porpoises navigate by sounds made by: contracting their muscles, slapping their fins against their bodies, using their vocal chords, blowing air out of their air holes.

69. The F-111 has the: agility of a fighter pursuit plane, capacity and endurance of a bomber, range of a transport, all of the above.

(70) The elimination of the screwworm fly was due to research sponsored by: the Cooperative Farmers Association, the Federal Government, a group of private chemical companies, the Florida Chamber of Commerce.

71. Units of measurement in the United States are defined by: the National Bureau of Standards, private industry, France, the Smithsonian Institution.

72. An example of technology is: determining the age of fossils, developing the artificial heart, experimenting with DNA, investigating cosmic ray sources.

73. Scientists are trying to unlock the energy of seawater because one gallon of seawater has as much energy potential as: 100 gallons of gasoline, 300 gallons of gasoline, 500 gallons of gasoline, 1,000 gallons of gasoline.

74. Phosphorite and manganese nodules have been discovered: on the moon, in the earth's core, on the ocean floor, in nebula.

75. Financial aid in education is available from: colleges and universities, the Federal Government, industrial organizations, all of the above.

76. Fluorescense microscopy is used in: tracing the communication patterns in skin language, studying the DNA molecule, determining the age of microfossils, tracking the movements of fish.

77. In the late 1890's, research in aerodynamics in this country was supported by: Grumman, Samuel P. Langley, the Federal Government, Institute of the Aeronautical Sciences.

78. Many scientists believe that a thorough understanding of DNA will lead to the: creation of many new chemical substances, elimination of hereditary diseases, utilization of salt water, verification of a theory of the universe.

79. Rockets and balloons are now being used to supplement information about the universe gained by various detecting devices used on earth because: the earth's atmosphere distorts and absorbs light and radio waves, the universe is rapidly expanding, the earth is somewhat pear-shaped, cosmic rays interfere with reception on earth.

80. The study of the earth's size, shape, mass, and gravitational field is called: geology, meteorology, geodesy, physics.

81. Scientific study of the earth from space vehicles has revealed that the earth has a tail which is formed mainly by the elongation of its: atmosphere, wind streams, magnetic field, dust clouds.

82. The mutations necessary in the development of a rot-resistant bean were accomplished by the technique of: phytochrome, eradication, radiation, crop spraying.

APPENDIX J

EXHIBIT-ONLY QUESTIONS

INSTRUCTIONS: Please check or circle the correct answer for each question. Answer all of the questions in the order they are given; do not skip around. Even if you are not sure of an answer, take a guess. There is no time limit on this test. Some of these questions may appear to have more than one right answer. However, if you have seen the exhibit, you should select the answer based on the information provided in the exhibit. Remember, answer every question.

1. The model of the skeleton found in the Rancho La Brea tar pits was: a clarkis pulchella, a pleistocene eagle, an atolla wyvillei, a porpoise
2. In the exhibit, the main objective of the chrysanthemum flower display was to show you the effect on plant growth of: seasons of the year, light and chemicals, cosmic inhibitors, high humidity refrigeration
3. This counter was activated by: the cosmic rays from space that passed through it, pushing a button on the side of the model, the microwave rays from a nearby energizing apparatus that passed through it, the rays from your body that passed through it.
4. The phrase NEW WAY TO TALK refers to the language of the: fingers, genetic code, eyes, skin
5. The "Coded Language of Life" refers to: polymerization, geodesy, genetics, fission
6. In the exhibit, the subatomic particle chart indicates: when each subatomic particle was identified, who discovered each subatomic particle, what each subatomic particle consists of, the number of electrons and neutrons in each subatomic particle.
7. In the exhibit, a model of DNA was shown in relation to: an ear of corn, a one-celled amoeba, a chrysanthemum, a baby
8. According to the exhibit, prehistoric conditions will be recreated in the laboratory in order to more thoroughly investigate: deoxyribonucleic acid, the age of the earth, fossil remains and imprints, the mutations of plant and animal species
9. What is the name of the exhibit? Vision of Science, Science and Technology, Man and Science, Vision of Man
10. According to the exhibit, Project Mohole is expected to greatly increase scientists' knowledge about: the shape of the earth, habits of fish, the ability of man to survive in the deep ocean, the age of the earth

11. In the next ten years, it is currently estimated that total school enrollment will increase by: 5 million, 10 million, 15 million, 20 million

⑫ In the exhibit, the evolutionary and steady-state theories are concerned with the: course of man's development, nature of atomic structure, history of the universe, age of the earth.

13. As shown in the exhibit model, the change from A to B above is intended to be a simplified illustration of the effect of: polymerization, phytochrome, DNA, extension

14. In order to use the space wrench correctly, you must: squeeze it, turn it clockwise, turn it counter-clockwise, move it up and down

15. This exhibit was paid for by: the American Association for the Advancement of Science, the Federal Government, General Electric Corporation, a group of private industries

⑯ In this exhibit, the model of the man amplifier was a: sound amplifier which man wears at his throat, powered exoskeleton device in which man sits, walking machine which man wears on his legs, device which man wears to regulate his heartbeat

17. The exhibit states that from 1776 to 1965 we have used as much fuel energy as we will use in the next: 5 years, 10 years, 20 years, 30 years

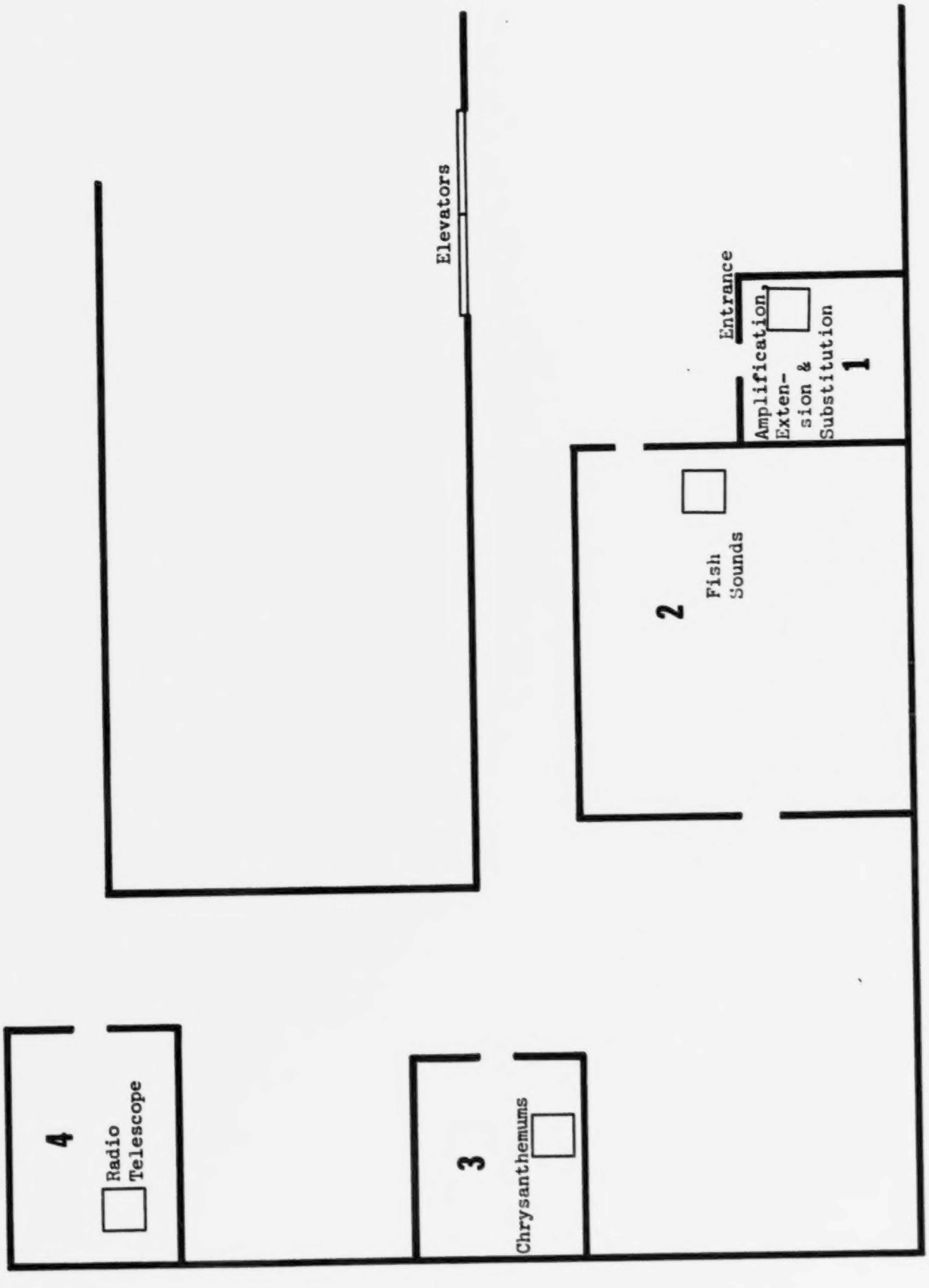
⑱ According to the exhibit, recent research shows that food may be safely preserved by passing it through a pool of: gibberellin, beta hydroxyethydrazine, polymers, cobalt 60 rods

19. The model in the exhibit that deals with amplification refers to: radio systems, muscle power, satellite communications, hearing

⑳ The original wood used in producing the irradiated "new wood" on the exhibit platform was: oak, cherry, walnut, pine

APPENDIX K

PHYSICAL LAYOUT OF THE MOCK-UP VARIATION


Elevators
Entrance
1
Amplification, Exten-sion & Substitution
2
Fish Sounds
3
Chrysanthemums
4
Radio Telescope

APPENDIX L

APPENDIX L

SUMMARY OF BACKGROUND QUESTIONNAIRE RESULTS

The graphic statistics presented in Table 37 allow several types of comparisons to be made across the subject groups tested at each exhibit site. These comparisons answer the following questions:

1. Were Los Angeles Casual Viewers similar to Chicago Casual Viewers in background?
2. Were Los Angeles study subjects similar to Chicago study subjects?
3. On what background characteristics did the casual viewer and study groups differ in Los Angeles?
4. On what background characteristics did the casual viewer and study groups differ in Chicago?

TABLE 37

Main Characteristics of Los Angeles and Chicago Viewer and Study Groups (All Experimental Groups Combined)

	(N=249) Los Angeles Casual Viewer	(N=197) Los Angeles Study	(N=364) Chicago Casual Viewer	(N=334) Chicago Study
Mean Age (years)	31	26	25	21
High School	17%	41%	33%	54%
College	13%	23%	23%	27%
Adult	70%	36%	44%	19%
Male	68%	52%	66%	54%
Female	32%	48%	34%	46%
Full-Time Students (Out of Total Age Group)	30%	64%	56%	81%
Adults Fully Employed (Out of all Adults)	74%	63%	84%	71%
Average Number High School Science Courses	1.4	2.2	2.2	2.6
Average Number College Science Courses	3.1	2.9	3.6	4.8
From Urban Area	48%	53%	52%	34%
From Suburban Area	44%	47%	35%	63%
From Rural Area	8%	0%	13%	3%
Median Highest Educational Level for Adults	13.1 yrs.	12.5 yrs.	12.8 yrs.	15.5 yrs.
Range for Adults	8-20 yrs.	9-18 yrs.	8-20 yrs.	8-20 yrs.

Comparison of casual viewer subjects. In the Los Angeles casual viewer groups there were proportionately more adults participating in the study than people of high school and college age. In Chicago, the age groups were more equally distributed. The Los Angeles adult concentration accounts for the differences between mean age and percentage of full time students at the two testing sites. The percentages of male-female adults fully employed and breakdowns of living area are quite similar. It appears that the Chicago casual viewer group had been exposed to more high school and college science courses, but this difference is not statistically significant. The median educational level for both adult groups was comparable and their years of schooling both ranged from an eighth grade education to the Ph.D. level. One can conclude that other than the concentration of adult age people in Los Angeles, the graphic statistics show similar background patterns for both of the Los Angeles and Chicago groups.

Comparison of study group subjects. The median age comparison and the percentage of full-time students reflect the greater concentration of high school students tested in Chicago. The large number of high school students tested in Chicago was deliberately planned in order to investigate the effect of socioeconomic levels upon the experimental variables. A larger percent of adults were tested in Los Angeles than in Chicago. The adult subjects in the Los Angeles testing were from a church group which had volunteered to participate in the study. The Chicago adult group consisted mainly of people who had been asked to participate at a local shopping center (near a university). The Chicago adult testing was at the mercy of those people who remembered and came to the museum. More Chicago adults were fully employed than in Los Angeles.

The male-female breakdowns was similar at both locations. The Chicago study group had more science courses in high school and college. The Los Angeles group lived in both urban and suburban areas, while a majority of Chicago participants lived in suburbia. Generally, the Chicago adults had more education than the Los Angeles group. The median educational level for Chicago was third year college, while Los Angeles people averaged at first year college. The range in number of years in school also supports this. The most schooling attained in Los Angeles subjects was the Masters Degree level, while the Ph.D. level was attained in Chicago.

In general, the Los Angeles and Chicago study groups were similar on male-female distribution. The Chicago group consisted of a higher concentration of high school students and a smaller percentage of adult participants. The Chicago subjects had a more scientific background as measured by number of science courses. The Chicago adults had more years of formal schooling.

Comparison of Los Angeles casual viewer and study subjects. A much larger percentage of adults were tested from the casual viewer group than from the study group. This percentage is reflected in total mean age difference, with casual viewers being generally older than the study group. Men were better represented in the casual group than in the study. There were more full time students in the study group, while more adults were fully employed in the casual groups. The casual group had on the average fewer high school science courses, but more college science courses. A small percentage of casual viewers were from a rural area while none of the study subjects lived in the country. The highest educational level for adults was similar, although the range in years of schooling was different.

The concentration of adults in the casual viewer group accounts for many of the discrepancies between the two groups. The experimental design for the Los Angeles testing was based on fairly equal age groups. One must conclude that the study and casual viewer groups are quite different on many of the measured background characteristics. However, because of the unbalanced age groupings, the degree of difference cannot be estimated.

Comparison of Chicago casual viewer and study groups. The Chicago casual viewer group also shows a higher percentage of adults than the study group, as would be expected because of the large concentration of high school students in the study group. The mean age for these two groups reflects the proportionate differences in the age groupings. More of the casual viewers were men, while the male-female breakdown was similar in the study group. The study group as a whole was better educated than the casual viewer group. They had more science courses in high school and college and the adults attained a higher educational level. A large proportion of the study group lived in a suburban area, while half of the casual viewer group were from the city.

The Chicago study and casual viewer groups look quite different on many of the background characteristics. Again, an estimate of the degree of difference is clouded by the disproportionate number of subjects within each age grouping.

APPENDIX M

APPENDIX M

INDIVIDUAL CELL FREQUENCIES AND MEANS FOR THE ANALYSES OF VARIANCE

Number of subjects in each condition.
Locale: Chicago
Analysis #1: Subjects = all non-casual subjects
Variables = experimental condition
sex
science background
age/educational level

INDIVIDUAL CELLS

Experimental Condition	Sex	Science Background			Non-science Background		
		HS	Coll	Adult	HS	Coll	Adult
Control	M	20	18	4	22	10	6
	F	8	17	6	18	7	21
Min	M	20	10	1	12	3	1
	F	14	3	2	12	3	9
Max	M	21	6	3	13	6	4
	F	11	3	2	10	4	4

GROUPS OF SUBJECTS

Experimental Condition	
Control	157
Min	90
Max	87
Sex	
Male	180
Female	154
Science Background	
Science	169
Non-science	165
Age/Educational Level	
High School	181
College	90
Adult	63

Means and standard deviations of criterion scores in each condition.
Locale: Chicago
Analysis #1: Subjects = all non-casual subjects
Variables = experimental condition
sex
science background
age/educational level
Criterion Score: Interest

CELL MEANS

Experimental Condition	Sex	Science Background			Non-science Background		
		HS	Coll	Adult	HS	Coll	Adult
Control	M	15.0	15.2	15.0	15.0	15.0	15.0
	F	13.9	15.0	15.0	15.0	15.0	15.0
Min	M	16.8	16.2	15.0	16.5	16.7	15.0
	F	17.6	16.3	17.0	16.1	16.0	15.3
Max	M	17.4	17.0	16.3	17.1	16.5	16.3
	F	17.2	18.0	16.0	16.9	17.5	15.5

OVERALL MEANS AND STANDARD DEVIATIONS:

	M	SD
Experimental Condition		
Control	15.0	0.8
Min	16.5	2.2
Max	17.0	1.4
Sex		
Male	16.0	1.3
Female	15.8	2.1
Science Background		
Science	16.1	1.6
Non-science	15.7	1.8
Age/Educational Level		
High School	16.2	1.6
College	15.7	1.4
Adult	15.3	2.2

Means and standard deviations of criterion scores in each condition.
Locale: Chicago
Analysis #1: Subjects = all non-casual subjects
Variables = experimental condition
sex
science background
age/educational level
Criterion Score: Open-end Concept

CELL MEANS

Experimental Condition	Sex	Science Background			Non-science Background		
		HS	Coll	Adult	HS	Coll	Adult
Control	M	7.7	10.6	7.3	6.3	8.8	7.5
	F	6.6	8.8	7.2	4.7	8.0	3.6
Min	M	9.7	13.3	14.0	8.7	14.7	11.0
	F	7.9	16.0	10.0	6.3	11.3	3.9
Max	M	11.3	13.8	16.3	11.2	16.7	10.8
	F	12.1	16.3	17.0	10.1	10.3	14.8

OVERALL MEANS AND STANDARD DEVIATIONS:

Experimental Condition	M	SD
Control	7.0	3.6
Min	9.1	4.1
Max	12.4	4.3
Sex		
Male	10.0	4.2
Female	7.8	4.5
Science Background		
Science	10.2	4.0
Non-science	7.7	4.6
Age/Educational Level		
High School	8.4	3.9
College	11.3	4.0
Adult	7.3	5.3

Means and standard deviations of criterion scores in each condition.
Locale: Chicago
Analysis #1: Subjects = all non-casual subjects
Variables = experimental condition
sex
science background
age/educational level
Criterion Score: Open-end Knowledge

CELL MEANS

Experimental Condition	Sex	Science Background			Non-science Background		
		HS	Coll	Adult	HS	Coll	Adult
Control	M	12.5	14.2	7.0	10.5	11.5	10.3
	F	10.8	12.6	9.0	7.4	11.9	6.1
Min	M	17.3	21.8	26.0	16.3	22.0	26.0
	F	14.6	20.0	20.0	11.5	18.3	7.0
Max	M	19.9	26.8	24.0	18.4	27.2	13.0
	F	16.2	26.0	24.5	16.6	17.0	22.8

OVERALL MEANS AND STANDARD DEVIATIONS:

Experimental Condition	M	SD
Condition	10.4	5.3
Min	16.0	6.7
Max	19.9	7.2
Sex		
Male	16.2	7.4
Female	12.3	6.9
Science Background		
Science	16.2	7.2
Non-science	12.6	7.2
Age/Educational Level		
High School	14.3	6.9
College	17.1	7.2
Adult	11.0	8.4

Means and standard deviations of criterion scores in each condition.
Locale: Chicago
Analysis #1: Subjects = all non-casual subjects
Variables = experimental condition
sex
science background
age/educational level
Criterion Score: Multiple Choice

CELL MEANS

Experimental Condition	Sex	Science Background			Non-science Background		
		HS	Coll	Adult	HS	Coll	Adult
Control	M	44.2	55.9	46.0	38.5	48.0	42.0
	F	40.5	45.6	45.5	34.3	42.9	33.2
Min	M	47.6	61.1	62.0	46.7	56.0	60.0
	F	42.0	55.3	47.0	35.7	55.0	27.0
Max	M	54.1	65.7	63.3	48.8	65.2	42.0
	F	46.4	66.3	64.0	45.3	55.0	60.5

OVERALL MEANS AND STANDARD DEVIATIONS:

	M	SD
Experimental Condition		
Control	42.3	10.9
Min	45.5	13.0
Max	53.6	13.5
Sex		
Male	49.9	12.7
Female	41.7	12.0
Science Background		
Science	50.1	12.2
Non-science	42.0	12.6
Age/Educational Level		
High School	43.8	11.6
College	54.2	10.5
Adult	41.2	15.1

Means and standard deviations of criterion scores in each condition.
Locale: Chicago
Analysis #1: Subjects = all non-casual subjects
Variables = experimental condition
sex
science background
age/educational level
Criterion Score: Exhibit Only

CELL MEANS

Experimental Condition	Sex	Science Background			Non-science Background		
		HS	Coll	Adult	HS	Coll	Adult
Control	M	8.5	10.6	7.0	7.7	9.5	7.8
	F	8.8	9.0	7.3	6.1	9.1	5.0
Min	M	12.1	14.5	10.0	11.8	14.0	16.0
	F	11.5	13.0	12.5	9.3	14.7	7.2
Max	M	13.0	17.0	16.3	12.2	14.8	11.3
	F	12.5	17.0	15.0	12.0	13.3	15.8

OVERALL MEANS AND STANDARD DEVIATIONS:

Experimental Condition	M	SD
Control	7.9	3.1
Min	11.6	3.5
Max	13.4	3.7
Sex		
Male	11.2	4.0
Female	9.4	4.1
Science Background		
Science	11.3	3.9
Non-science	9.3	4.0
Age/Educational Level		
High School	10.3	3.8
College	11.9	3.7
Adult	8.4	4.6

Number of subjects in each condition.
Locale: Chicago
Analysis #2: Subjects = all high school non-casual subjects
Variables = experimental condition
sex
science background
socioeconomic level

INDIVIDUAL CELLS

Experimental Condition	Sex	Science Background			Non-science Background		
		High	Med	Low	High	Med	Low
Control	M	5	4	11	11	3	8
	F	1	2	5	10	2	6
Min	M	6	2	12	3	5	4
	F	1	4	9	8	1	3
Max	M	4	7	10	3	1	9
	F	2	3	6	5	1	4

GROUPS OF SUBJECTS

Experimental Condition	
Control	68
Min	58
Max	55
Sex	
Male	108
Female	73
Science Background	
Science	94
Non-science	87
Socioeconomic Level	
High	59
Medium	35
Low	87

Means and standard deviations of criterion scores in each condition.
Locale: Chicago
Analysis #2: Subjects = all high school non-casual subjects
Variables = experimental condition
sex
science background
socioeconomic level
Criterion Score: Interest

CELL MEANS

Experimental Condition	Sex	Science Background			Non-science Background		
		High	Med	Low	High	Med	Low
Control	M	15.0	15.0	15.0	15.0	15.0	15.0
	F	15.0	15.0	13.2	15.0	15.0	15.0
Min	M	16.7	17.0	16.8	15.7	16.2	17.5
	F	19.0	16.8	17.8	16.0	15.0	16.7
Max	M	17.3	17.7	17.3	17.0	17.0	17.1
	F	17.0	18.3	16.7	16.8	15.0	17.5

OVERALL MEANS AND STANDARD DEVIATIONS:

Experimental Condition	M	SD
Control	14.9	1.1
Min	16.8	1.3
Max	17.2	1.4
Sex		
Male	16.2	1.4
Female	16.1	1.9
Science Background		
Science	16.5	1.8
Non-science	15.9	1.3
Socioeconomic Level		
High	15.9	1.4
Medium	16.4	1.5
Low	16.3	1.8

Means and standard deviations of criterion scores in each condition.
Locale: Chicago
Analysis #2: Subjects = all high school non-casual subjects
Variables = experimental condition
sex
science background
socioeconomic level
Criterion Score: Open-end Concept

CELL MEANS

Experimental Condition	Sex	Science Background			Non-science Background		
		High	Med	Low	High	Med	Low
Control	M	9.2	8.3	6.7	6.7	5.0	6.1
	F	8.0	8.0	5.8	4.3	7.0	4.5
Min	M	11.3	10.0	8.8	8.3	8.0	9.8
	F	10.0	11.0	6.2	5.4	10.0	7.3
Max	M	12.5	12.1	10.3	17.7	14.0	8.7
	F	16.0	12.7	10.5	11.0	13.0	8.3

OVERALL MEANS AND STANDARD DEVIATIONS:

Experimental Condition	M	SD
Control	6.3	3.0
Min	8.3	3.1
Max	11.2	4.1
Sex		
Male	9.0	3.9
Female	7.6	3.9
Science Background		
Science	9.4	3.8
Non-science	7.4	3.9
Socioeconomic Level		
High	8.6	4.5
Medium	9.8	3.3
Low	7.8	3.6

Means and standard deviations of criterion scores in each condition.
Locale: Chicago
Analysis #2: Subjects = all high school non-casual subjects
Variables = experimental condition
sex
science background
socioeconomic level
Criterion Score: Open-end Knowledge

CELL MEANS

Experimental Condition	Sex	Science Background			Non-science Background		
		High	Med	Low	High	Med	Low
Control	M	15.8	12.8	10.8	11.5	7.7	10.1
	F	10.0	11.0	10.8	6.8	10.5	7.3
Min	M	22.5	19.5	14.3	10.3	17.8	18.8
	F	13.0	20.3	12.3	10.9	20.0	10.3
Max	M	22.0	21.3	18.0	25.3	18.0	16.1
	F	22.5	16.7	13.8	19.0	19.0	13.0

OVERALL MEANS AND STANDARD DEVIATIONS:

Experimental Condition	M	SD
Control	10.3	4.6
Min	15.2	6.0
Max	18.2	6.7
Sex		
Male	15.5	6.5
Female	12.4	6.2
Science Background		
Science	15.7	6.8
Non-science	12.7	6.0
Socioeconomic Level		
High	14.5	7.0
Medium	16.6	5.5
Low	13.2	6.5

Means and standard deviations of criterion scores in each condition.
Locale: Chicago
Analysis #2: Subjects = all high school non-casual subjects
Variables = experimental condition
sex
science background
socioeconomic level
Criterion Score: Multiple Choice

CELL MEANS

Experimental Condition	Sex	Science Background			Non-science Background		
		High	Med	Low	High	Med	Low
Control	M	53.4	50.0	37.8	40.1	35.3	37.6
	F	40.0	49.0	37.2	34.6	46.0	29.8
Min	M	55.2	58.0	42.0	41.0	48.0	49.3
	F	46.0	51.0	37.6	35.3	47.0	33.0
Max	M	53.3	58.7	51.3	63.7	59.0	42.7
	F	58.0	47.3	42.0	48.2	62.0	37.5

OVERALL MEANS AND STANDARD DEVIATIONS:

Experimental Condition	M	SD
Control	39.3	9.5
Min	43.6	10.5
Max	49.7	12.7
Sex		
Male	46.4	11.7
Female	40.0	10.4
Science Background		
Science	46.7	11.7
Non-science	40.7	10.7
Socioeconomic Level		
High	44.7	12.0
Medium	50.8	10.0
Low	40.4	10.7

Means and standard deviations of criterion scores in each condition.
Locale: Chicago
Analysis #2: Subjects = all high school non-casual subjects
Variables = experimental condition
sex
science background
socioeconomic level
Criterion Score: Exhibit Only

CELL MEANS

Experimental Condition	Sex	Science Background			Non-science Background		
		High	Med	Low	High	Med	Low
Control	M	9.6	9.5	7.5	8.4	6.7	7.3
	F	10.0	10.5	7.8	5.9	7.0	6.2
Min	M	13.8	15.0	10.8	11.3	11.4	12.5
	F	15.0	15.0	9.6	9.4	11.0	8.3
Max	M	12.3	15.1	11.7	14.3	11.0	11.7
	F	16.5	14.0	10.3	12.2	16.0	10.8

OVERALL MEANS AND STANDARD DEVIATIONS:

	M	SD
Experimental Condition		
Control	7.6	2.5
Min	11.3	3.1
Max	12.5	3.9
Sex		
Male	10.7	3.9
Female	9.7	3.6
Science Background		
Science	11.2	3.9
Non-science	9.3	3.5
Socioeconomic Level		
High	10.2	3.9
Medium	12.2	3.8
Low	9.6	3.5

NUMBER OF SUBJECTS IN EACH CONDITION

Locale: Chicago
Analysis #3: Subjects = all subjects
Variables = experimental condition
science background
age/educational level

INDIVIDUAL CELLS

Experimental Condition	Science Background			Non-science Background		
	HS	Coll	Adult	HS	Coll	Adult
Casual-post	1	4	3	11	1	6
Control	28	35	10	40	17	27
Min	34	13	3	24	6	10
Max	32	9	5	23	10	8

GROUPS OF SUBJECTS

Experimental Condition	
Casual-post	26
Control	157
Min	90
Max	87
Science Background	
Science	177
Non-science	183
Age/Educational Level	
High School	193
College	95
Adult	72

Means and standard deviations of criterion scores in each condition.
Locale: Chicago
Analysis #3: Subjects = all subjects
Variables = experimental condition
science background
age/educational level
Criterion Score: Interest

CELL MEANS

Experimental Condition	Science Background			Non-science Background		
	HS	Coll	Adult	HS	Coll	Adult
Casual Post	2.0	6.0	4.0	4.8	5.0	3.5
Control	14.7	15.1	15.0	15.0	15.0	15.0
Min	17.1	16.2	16.3	16.3	16.3	15.3
Max	17.3	17.3	16.2	17.0	16.9	15.9

OVERALL MEANS AND STANDARD DEVIATIONS:

Experimental Condition	M	SD
Casual Post	4.5	2.0
Control	15.0	0.8
Min	16.5	2.2
Max	17.0	1.4
Science Background		
Science	15.6	2.9
Non-science	14.6	3.8
Age/Educational Level		
High School	15.5	3.2
College	15.2	2.6
Adult	13.9	4.5

Means and standard deviations of criterion scores in each condition.
Locale: Chicago
Analysis #3: Subjects = all subjects
Variables = experimental condition
science background
age/educational level
Criterion Score: Open-end Concept

CELL MEANS

Experimental Condition	Science Background			Non-science Background		
	HS	Coll	Adult	HS	Coll	Adult
Casual Post	6.0	8.5	6.3	5.2	6.0	5.5
Control	7.4	9.7	7.2	5.6	8.5	4.4
Min	8.9	13.9	11.3	7.5	13.0	4.6
Max	11.6	14.7	16.6	10.7	14.1	12.8

OVERALL MEANS AND STANDARD DEVIATIONS:

Experimental Condition	M	SD
Casual Post	6.0	2.9
Control	7.0	3.6
Min	9.1	4.1
Max	12.4	4.3
Science Background		
Science	10.1	4.0
Non-Science	7.5	4.5
Age/Educational Level		
High School	8.2	3.9
College	11.1	4.1
Adult	7.1	5.0

Means and standard deviations of criterion scores in each condition.
Locale: Chicago
Analysis #3: Subjects = all subjects
Variables = experimental condition
science background
age/educational level
Criterion Score: Open-end Knowledge

CELL MEANS

Experimental Condition	Science Background			Non-science Background		
	HS	Coll	Adult	HS	Coll	Adult
Casual Post	11.0	12.3	14.0	14.3	12.0	8.0
Control	12.0	13.5	8.2	9.1	11.6	7.1
Min	16.2	21.4	22.0	13.9	20.2	8.9
Max	18.6	26.6	24.2	17.6	23.1	17.9

OVERALL MEANS AND STANDARD DEVIATIONS:

	M	SD
Experimental Condition		
Casual Post	12.3	8.5
Control	10.4	5.3
Min	16.0	6.7
Max	19.9	7.2
Science Background		
Science	16.0	7.1
Non-science	12.5	7.5
Age/Educational Level		
High School	14.2	7.0
College	16.8	7.1
Adult	10.9	8.0

Means and standard deviations of criterion scores in each condition.
Locale: Chicago
Analysis #3: Subjects = all subjects
Variables = experimental condition
science background
age/educational level
Criterion Score: Multiple Choice

CELL MEANS

Experimental Condition	Science Background			Non-science Background		
	HS	Coll	Adult	HS	Coll	Adult
Casual Post	65.0	65.5	62.0	61.4	51.0	56.7
Control	43.1	50.9	45.7	36.6	45.9	35.1
Min	45.3	59.8	52.0	41.2	55.5	30.3
Max	51.5	65.9	63.6	47.3	61.1	51.3

OVERALL MEANS AND STANDARD DEVIATIONS:

Experimental Condition	M	SD
Casual Post	60.7	9.8
Control	42.3	10.9
Min	45.5	13.0
Max	53.6	13.5
Science Background		
Science	50.8	12.3
Non-science	43.7	13.4
Age/Educational Level		
High School	44.9	12.4
College	54.6	10.6
Adult	43.3	15.4

Means and standard deviations of criterion scores in each condition.
Locale: Chicago
Analysis #3: Subjects = all subjects
Variables = experimental condition
science background
age/educational level
Criterion Score: Exhibit Only

CELL MEANS

Experimental Condition	Science Background			Non-science Background		
	HS	Coll	Adult	HS	Coll	Adult
Casual Post	5.0	8.3	8.7	10.1	8.0	9.8
Control	8.5	9.8	7.2	7.0	9.4	5.6
Min	11.9	14.2	11.7	10.5	14.3	8.1
Max	12.8	17.0	15.8	12.1	14.2	13.5

OVERALL MEANS AND STANDARD DEVIATIONS:

Experimental Condition	M	SD
Casual Post	9.3	4.0
Control	7.9	3.1
Min	11.6	3.5
Max	13.4	3.7
Science Background		
Science	11.2	3.9
Non-science	9.4	4.1
Age/Educational Level		
High School	10.2	3.9
College	11.7	3.8
Adult	8.5	4.4